The Pennsylvania State University

The Graduate School

Department of Statistics

NONPARAMETRIC IMPUTATION AND (MID-) RANK TESTS FOR MIXED EFFECTS MODELS WITH MISSING DATA

A Thesis in

Statistics

by

Efi S. Antoniou

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2004

The thesis of Efi S. Antoniou was reviewed and approved* by the following:

Michael G. Akritas
Professor of Statistics
Thesis Adviser
Chair of Committee

Thomas P. Hettmansperger
Professor of Statistics

Runze Li
Assistant Professor of Statistics

Linda M. Collins
Professor of Human Development and Family Studies

James L. Rosenberger
Professor of Statistics
Head of the Department of Statistics

* Signatures are on file in the Graduate School.

Abstract

The first part of this thesis deals with factorial designs where each subject is observed at several time points with part of the data missing. A nonparametric approach for estimating the marginal cumulative distribution function at each time point is proposed and used to test for factor effects and interactions. Estimation uses more general and flexible donor sets which leads to a new type of nonparametric imputation. In particular, the donor sets allow use of univariate kernel methods even with higher dimensional data, avoiding thus the curse of dimensionality. The classical missing at random assumption is not tailored for the present nonparametric analysis. The notion of the missingness conditionally at random Comparisons with ML indicate that the proposed method fares well when the data are normal and homoscedastic, and outperforms it in other cases.

The second part of this thesis considers testing for covariate-adjusted main effects and interactions in the context of the fully nonparametric ANCOVA model. The test procedures of Akritas, Arnold and Du (2000) are based on consistent estimation of the conditional distributions and as such they involve the cumbersome task of bandwidth determination. The proposed methodology does not require such consistent estimation. Asymptotic theory and numerical results, indicate that nearest neighbor windows of fixed (small) size perform well. This makes the applicability of the fully nonparametric methodology in real-life situations easily feasible.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am deeply indebted to many people for their strong support during my graduate studies. First, I would like to thank my advisor Dr. Michael G. Akritas for being an exceptional advisor and friend during my graduate studies. His valuable feedback, guidance, patience and support have been instrumental in designing and writing this study. I would also like to thank all my committee members, Dr. Linda M. Collins, Dr. Thomas P. Hettmansperger and Dr. Runze Li, for their interesting questions and comments. Also thanks to Dr. Wayne Osgood for providing me with the sociology data.

I also want to express my deep gratitude to the faculty and graduate students at Penn State Statistics Department. Special thanks to Lan Wang for working with me and for her valuable discussions, Haiyan Wang for being a close friend during all the years I spent here and Nadia Pashiardi for being very supportive during this past year. Also, thanks to all my friends that have already left and especially Ryan Elmore, Ofer Harel, James McDermott and Mihaella Ugargovici.

I am truly grateful to my parents, Savvas and Eleni for their endless love, patience and encouragement, and my sisters, Polina and Constantina, who supported me in each step I took towards my personal and professional growth.

Finally, I would like to thank all my friends in State College and in Cyprus and especially Paris Katselas for always being there for me.

Chapter 1

INTRODUCTION

1.1 Missing Data Analysis Background

In a wide range of applied studies it is almost inevitable missing values to appear in the data sets. Data with missing values causes difficulties to scientific research because of ambiguities regarding the appropriate method of analysis. Depending on the design of the study, missing information can be partially recovered from the observed data. For doing so, we have to consider whether or not missingness is related to the data. Therefore, missing data analyses are based on assumptions about the processes that create the missing values. The simplest such assumption is that there is no relation between data and missingness, i.e. that missingness is completely at random (MCAR). It has been found, however, that MCAR is often violated in practice. Moreover, methods that rely on MCAR give biased results when this assumption does not hold. Thus, the recent literature deals with more general assumptions. In the following we review such assumptions and procedures considering parametric and nonparametric approaches separately.

1.1.1 Parametric Approaches

From the parametric point of view, there is a well developed theory for missing data problems which is nicely described in Schafer (1997). The two main classes of parametric procedures, maximum likelihood (ML) and multiple imputation (MI) are built upon the assumption of *ignorability* (Little and Rubin, 1987), which states that the data are missing at random (MAR) (Rubin, 1976) and that the parameter space of the full data model and the parameter space of the missingness mechanism are distinct. To describe the MAR assumption let $\mathbf{Y}$ denote the random variables we wish to observe, and let $\mathbf{\Delta}$ denote corresponding variables indicating whether or not each variable in $\mathbf{Y}$ is observed. Moreover, let $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$ be the observed and missing parts of $\mathbf{Y}$ respectively. Then, we have MAR if for any value of the parameter ϕ of the missingness mechanism (not necessarily a scalar),

$$P(\mathbf{\Delta} = \boldsymbol{\delta}|\mathbf{Y}_{obs} = \mathbf{y}_{obs}, \mathbf{Y}_{mis}, \phi) = P(\mathbf{\Delta} = \boldsymbol{\delta}|\mathbf{Y}_{obs} = \mathbf{y}_{obs}, \phi), \qquad (1.1.1)$$

where $\boldsymbol{\delta}$ and $\mathbf{y}_{obs}$ are realizations of $\mathbf{\Delta}$ and $\mathbf{Y}_{obs}$ in the sample. Note that at the theoretical level, (1.1.1) is not required to hold for values of $\mathbf{\Delta}$ and $\mathbf{Y}_{obs}$ other than the observed ones. In words, MAR allows the probabilities of missingness to depend on the observed data but not on the missing data. The simpler MCAR assumption is a special case of MAR obtained by assuming that these probabilities do not depend on the observed data either. Denoting the parameter of the data model with θ (not necessarily a scalar),

$$\begin{aligned} P(\mathbf{\Delta}, \mathbf{Y}_{obs}|\ \theta, \phi) &= \int P(\mathbf{\Delta}, \mathbf{Y}|\ \theta, \phi) d\mathbf{Y}_{mis} = \int P(\mathbf{\Delta}|\ \mathbf{Y}, \phi) P(\mathbf{Y}|\ \theta) d\mathbf{Y}_{mis} \\ &= P(\mathbf{\Delta}|\mathbf{Y}_{obs}, \phi) \int P(\mathbf{Y}|\ \theta) d\mathbf{Y}_{mis} = P(\mathbf{\Delta}|\mathbf{Y}_{obs}, \phi) P(\mathbf{Y}_{obs}|\ \theta), \end{aligned}$$

where the third equality holds under the MAR assumption. If the parameter spaces of θ and ϕ are separable, this factorization of the observed data ($\mathbf{Y}$ and $\mathbf{\Delta}$) likelihood implies that likelihood based inference for θ can be performed ignoring the missing data mechanism since

$$L(\theta|\mathbf{Y}_{obs}) \propto P(\mathbf{Y}_{obs}|\theta).$$

Little and Rubin refers to $P(\mathbf{Y}_{obs}|\theta)$ as the likelihood ignoring the missing-data mechanism, and Schafer (1997) as the observed-data likelihood. Rubin (1976) demonstrated that (1.1.1) is the weakest general condition under which we may ignore the missingness mechanism in likelihood or Bayesian analysis.

As in the analysis of complete data, ML procedures have appealing theoretical properties in incomplete data cases as well. Little and Rubin (1987) provide full details for this analysis provided the parametric model is correct. The form of the observed data likelihood, is complicated, and except special cases, there are no close form expressions for the ML estimates of θ. Hence application of the ML method require specialized algorithms. Dempster, Laird and Rubin (1977) introduced the EM algorithm to solve this type of optimization problems. Little and Rubin (1987) and Schafer (1997) give nice overviews of the EM algorithm. Normal based ML procedure with unstructured covariance matrix is implemented in several programs; e.g. SPSS (Norusis, 2000), EMCOV (Graham and Hofer, 1991), NORM (Schafer, 1999). ML is also available for normal models with structured or unstructured covariance matrices; e.g. SAS procedure PROC MIXED (Little et. al., 1996), nlme package in S-plus and R (Pinheiro and Bates, 2000). These procedures are designed to be used for unbalanced repeated measures data sets, where unbalanced refers also to units being measured at different time points. In particular the documentation of these procedures does not mention missing values, but missingness can be considered as imbalance occurring by design. Collins, Schafer and Kam (2001) pointed out that ML analysis under the MAR assumption for normal models with structured covariance matrix (but same covariance matrix in all groups) can be carried out by ignoring all missing values and feeding the observed data in the SAS PROC MIXED procedure. Equivalent analysis can be performed using the nlme package in S-plus or R.

Multiple imputation is an alternative parametric approach to missing data problems proposed by Rubin (1987). The main difference from the ML methods is that missingness is treated in a separate stage before the analysis stage. In MI, each missing value is replaced by a list of $M > 1$ values, producing M complete data sets. Then, each of the data sets is analyzed using standard complete data sets methods, and finally, the M sets of results are combined based on some rules. In order to generate the imputations first one needs to specify, in addition to the parametric model for the complete data, a prior distribution for the parameters, and then use Bayesian arguments to simulate independent draws from the distribution of $\mathbf{Y}_{mis}$ given Y_{obs}. In most of the cases this is carried out using MCMC computational techniques.

In the simplest case of a single scalar parameter Rubin (1987) proposed using the average of the M point estimators and rules for combing their standard errors. When interest lies in higher dimensional parameters, as in the case in the analysis of factorial designs which we consider, the MI approach is not so straightforward and the available solutions not very satisfactory. Li, Raghunathan and Rubin (1991), who considered this problem, proposed again the use of the average of the M point estimators but their proposal for combining the covariance matrices works well only under the stringent condition that the fractions of missing information for all components of the higher-dimensional parameter are equal (see also Schafer, 1997, pp. 113, 114). The expression 'works well' in the above statements refers to the use of the covariance matrix for constructing Wald-type test statistics. As a way out of this difficulty, Li, Meng, Raghunathan and Rubin (1991) considered approaches to testing based on direct combination of either complete data p-values or test statistics. However, they pointed out that the results are not very accurate and thus recommended that this approach be used only as an exploratory tool; in particular, they recommend looking at half and double the p-value obtained before reaching a conclusion. We are not aware of subsequent work improving the state of the art of MI with higher-dimensional parameters.

There are many computer programs available for MI: For imputations under the multivariate normal with unstructured covariance matrix (but same in all groups) see NORM, which is also implemented in SAS PROC MI, in S-plus Amelia library (Schimert et. al., 2001) and in R norm package. The S-plus function PAN (Schafer, 2001) also implements normal based imputations with a structured covariance matrix. For multivariate categorical data and mixed data sets (continuous and categorical variables) Schafer's CAT and MIX S-plus libraries are implement in the S-plus missing-data module (Schimert et. al., 2001).

Collins, Schafer and Kam (2001) present a nice comparison of the ML and MI procedures when interest lies in a scalar paramter. They pointed out that when the same model is used for imputation and analysis, the MI procedures yields similar results to the ones of ML procedure using the same model.

It should also be mentioned that recent research activity considers the case that missingness is not at random (MNAR), i.e. the probabilities of missingness depend on both $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$. This type of missingness has applications in several clinical trials were the drop out is closely related to the variable being measured. There are two main ways to perform data analysis assuming MNAR, selection models and pattern mixture models. In both of these analyses one must specify the missingness mechanism which often is a complicated problem. Rubin (1995) is reviewing selection models and pattern mixture models for longitudinal studies. Recently, Tang, Little and Raghunathan (2003) considered multivariate regression analysis with nonignorable missingness. Their method does not require specification of the form of the nonresponse mechanism.

1.1.2 Nonparametric Approaches

There are two veins of nonparametric approaches with missing data, one which aims to extend complete data (mid-)rank testing procedures and another which aims to relax the parametric assumptions in the parametric imputation methods through the use of smoothing techniques.

In the first vein belong the complete and available cases methods (see Brunner, Munzel and Puri, 1999) which, however, are valid under the rather restrictive MCAR assumption. In the second vein there is a more extended literature. Titterington and Mill (1983) introduced kernel methods for imputation of missing values. To briefly describe this method let $\mathbf{X} = (X_1, \ldots, X_k)$ denote the variables one would like to observe. They considered a nonparametric estimation of the joint distribution of $\mathbf{X}$ using kernel methods to recover the missing information. For every case with missing values the complete cases are playing the role of the 'donor set' since the smoothing is done using the observed values of the incomplete case and the corresponding values of these variables in the complete cases. This theory is valid under a missingness assumption stronger than MAR, which coincides with the MAR only in the simpler case where $\mathbf{X} = (X_1, X_2)$ and $\mathbf{X}_1$ is always observed. Chen (1994), considered cases where $X_1, \ldots, X_{k-1}$ are fully observed, and $X_k = Y$ is missing under the *strongly ignorable MAR* proposed by Rosenbaum and Rubin (1983) (see Remark 1.1.1). They derived nonparametric estimation of mean functionals of Y based on nonparametric estimation of the mean of the conditional density of Y given the X's. Cheng and

Chum (1996) complement this theory by estimating the marginal cdf of Y.

Remark 1.1.1. *The strongly ignorable MAR assumption states that the missingness mechanism is conditionally independent from the variables that are subject to missingness given the variables that are always observed. For example, consider a data set consisted by independent replications of the random vector* $\mathbf{X} = (X_1, \ldots, X_k)$*. If* $X_1, \ldots, X_m$ *are always observed and* $X_{m+1}, \ldots, X_k$ *are subject to missingness, then*

$$P(\boldsymbol{\Delta}|\mathbf{X}) = P(\boldsymbol{\Delta}|X_1, \ldots, X_k) = P(\boldsymbol{\Delta}|X_1, \ldots, X_m),$$

where $\boldsymbol{\Delta}$ *is the missingness indicator random vector* $(\Delta_1, \ldots, \Delta_k) = (1, \ldots, 1, \Delta_{m+1}, \ldots, \Delta_k)$*. In general, nonparametric methods are derived assuming strongly ignorable MAR since in contrast to the MAR assumption, is a probabilistic rule. However, a lot of the authors make no distinction between the two assumptions.*

Titterington and Sedransk (1989) used kernel density estimation in combination with a nonparametric bootstrap for imputing missing values. This method is considering multiple imputations in contrast to the single imputation methodology proposed in the aforementioned nonparametric procedures. This method does not directly account for the relationship between variables since the missing values for one variable are imputed using either bootstrap or smoothed bootstrap involving its observed values. Efron (1994) developed a nonparametric analogue to the approximate Bayesian bootstrap idea proposed by Rubin (1987). Efron's method handles only nominal categorical variables and it does not use any smoothing techniques. Aerts et. al. (2002) considered nonparametric smoothing methods to obtain MI estimators in a non-Bayesian framework in the sense of Efron's (1994) bootstrap. Their method deals with cases where some of the variables are fully observed and they assume strongly ignorable MAR. In this paper they consider the simpler case where there is one fully observed covariate and a univariate response incompletely observed and of interest is a marginal parameter the response. The extension of this method to cases with more than one fully observed covariates demands the use of high dimension kernels which results to curse of dimensionality problems. In the simpler bivariate situation, for every case with missing response the donor set is the complete cases set and this is going to be true in the multi-covariate case with a univariate response.

Even though the idea of nonparametric imputation is not new, to the best of our knowledge the use of this idea to construct rank or other type test statistics is first proposed in Akritas, Osgood and Kuha (2002). In fact, until recently, there was a common perception that it is not possible to implement nonparametric testing procedures without assuming MCAR. Akritas et. al. (2002) have shown that it is possible to relax this assumption in the case where subjects are observed at two follow up times when no other factors are involved. However, the classical MAR assumption are not suitable for the proposed nonparametric analysis, and thus Akritas et. al. (2002) introduced a new missingness assumption which we call *missing conditionally at random* (MCR). The MCR assumption is the weakest condition needed for nonparametric analysis, i.e. nonparametric imputation and estimation of the underlying population distributions. The MCR assumption and the MAR assumption are qualitatively different and do not imply each other. In this thesis the ideas of Akritas et. al. (2002) are extended to factorial designs with repeated measurements when some of the data is missing.

The main contribution in the literature of nonparametric imputation is relaxing the assumption of strong ignorability. This is achieved though a representation of the marginal distributions, which allowed the use of more general and flexible donor sets in our nonparametric imputation.

1.2 Fully Nonparametric Methods for Complete Data

The models for different analyses presented in this thesis are based on fully nonparametric procedures that were developed for complete data. In this section we are giving a short introduction to some fully non-parametric models relevant to methods proposed in the later chapters.

1.2.1 Analysis of Repeated Measure Designs

Akritas and Arnold (1994) first introduced fully nonparametric analysis for multivariate repeated measures designs. For simplicity, in the case of a two-way crossed design the nonparametric model only specifies that $Y_{ijk} \sim F_{ij}$ for all i and j. Denoting

$$F_{ij}(y) = \frac{1}{2}[P(Y_{ij} < y) + P(Y_{ij} \leq y)],$$

and defining accordingly the empirical distribution function, allows a unified formulation of (mid-) rank statistics and test procedures for all ordinal data. This general model is fully nonparametric, does not require homoscedasticity, and it can treat continuous and ordinal data. Basically, this fully nonparametric approach replaces the vector of expectation, which is used in the theory of linear models, by the vector of marginal distribution functions. Thus, main effects and interactions are defined by a linear decomposition of the distribution functions, similar to that of the means. Similar to the classical ANOVA decomposition F_{ij} can be uniquely decomposed to

$$F_{ij}(x) = M(x) + A_i(x) + B_j(x) + C_{ij}(x), \tag{1.2.1}$$

where $\sum_i A_i = \sum_j B_j = 0$, $\sum_i C_{ij} = 0$, for all j and $\sum_j C_{ij} = 0$ for all j. In fact, $M(x) = \bar{F}_{..}(x)$, $A_i(x) = \bar{F}_{i.}(x) - \bar{F}_{..}(x)$, $B_j(x) = \bar{F}_{j.}(x) - \bar{F}_{..}(x)$, and $C_{ij}(x) = F_{ij}(x) - \bar{F}_{i.}(x) - \bar{F}_{j.}(x) + \bar{F}_{..}(x)$.

Based on this decomposition the following hypotheses of no main effects and no interaction are formed

$$\begin{aligned} H_0(A) &: A_i(x) = 0,\ \forall i, \\ H_0(B) &: B_j(x) = 0,\ \forall j, \\ H_0(AB) &: C_{ij} = 0,\ \forall\ i,\ j. \end{aligned}$$

The test statistics proposed to test these hypotheses are the rank transform versions of the classical statistics for testing hypotheses in repeated measures designs. Note that even thought the rank transform method in some cases is inappropriate for testing parametric hypotheses (Akritas, 1990) is always valid for testing the nonparametric hypotheses. This nonparametric hypotheses and the corresponding tests statistics are invariant under monotone transformations of the response which is an appealing feature especially in the case of ordinal data. Also it should be noted that the nonparametric effects imply their parametric counterparts.

Throughout the last decade there were several developments of this nonparametric analysis to different directions. This work was based on the main idea of the nonparametric model to only specify the distribution function of each combination of values of the explanatory variables. The rank transform versions of the classical statistics used in Akritas and Arnold (1994) can handle only homoscedastic data, in the later papers this restriction is relaxed by using Walt-type (mid)-rank statistics. Akritas, Arnold and Brunner (1997) developed nonparametric analysis for factorial designs with independent data, and Akritas and Brunner (1997) presented a general framework dealing with arbitrary mixed models.

1.2.2 Analysis of Covariance

Akritas, Arnold and Du (2000) introduced a nonparametric ANCOVA model with one covariate. This model only assumes conditional independence of the response given the covariate with conditional distribution function. For example, consider a one-way ANCOVA model with a groups. Suppose we observe (Y_{ij}, X_{ij}), $i = 1, \ldots, a$, j=1,...,n_i, where Y_{ij} and X_{ij} represent the response and the covariate in the jth observation of the ith group. With this notation, and denoting

$$F_i(y|x) = F_{ix}(y) = \frac{1}{2}\left[P(Y_{ij} \leq y|X_{ij} = x) + P(Y_{ij} < y|X_{ij} = x)\right]. \quad (1.2.2)$$

the fully nonparametric model assumes only that

$$Y_{ij}|X_{ij} = x \sim F_{ix} \ , \quad (1.2.3)$$

i.e., that conditionally on $X_{ij} = x$, Y_{ij} has distribution function that depends on i and x. Note that model (1.2.3) does not specify how the response distribution changes when the levels, or covariate value changes, and does not assume continuity of the conditional distributions. Thus it is completely nonparametric (also nonlinear and non-additive). In order to define effects and hypotheses in this nonparametric context, let $G\left(x\right)$ be the distribution of the covariate pooled over the whole population and set

$$\overline{F}_{i\cdot}\left(y\right) = \int F_{ix}\left(y\right) d\ G\left(x\right), \text{ and } \overline{F}_{\cdot x}\left(y\right) = \frac{1}{a}\sum_{i=1}^{a} F_{ix}\left(y\right) \ .$$

Using this notation, the conditional distribution function is uniquely decomposed as follows

$$F_{ix}(y) = M(y) + A_i(x) + D_x(y) + C_{ix}(y),$$

where $M(y) = a^{-1}\sum_{i=1}^{a}\int F_{ix}(y)dG(x)$, $A_i(y) = \overline{F}_{i\cdot}(y) - M(y)$, $D_x(y) = \overline{F}_{\cdot x}(y) - M(y)$, $C_{ix}(y) = F_{ix}(y) - M(y) - A_i(y) - D_x(y)$. In this decomposition, A_i are the covariate-adjusted nonparametric main effects of the factor, D_x is the nonparametric covariate effect and C_{ix} the nonparametric interaction between the factor and the covariate. Akritas, Arnold and Du (2000) proposed a test for the hypothesis of no main factor effect, $A_i = 0$, for all i.

Their approach relies on consistent estimation (using kernel methods) of the conditional distribution functions F_{ix}. By its nature, application of this approach requires determination of the window bandwidth which is particularly cumbersome in this context. (In Akritas, Arnold and Du (2000) the bandwidth choice was based on resampling from the data in a way that imitates the null hypothesis.) Akritas, Antoniou, and Wang (2004) considered an alternative test procedure for the same hypotheses. The main novelty of the new procedure is that it does not require consistent estimation of the F_{ix} and the estimator is obtained using nearest neighbor windows of *fixed* size.

Tsangari and Akritas (2003) generalized the methodology of Akritas, Arnold and Du (2000) to include ANCOVA designs with two and three covariates. However, curse of dimensionality effects prevent further generalization to more covariates. Since the methodology presented in Akritas, Antoniou, and Wang (2004) uses fixed window sizes, it will be less affected by the curse of dimensionality and thus holds the promise of extendability to more than three covariates. This will be investigated in a forthcoming paper.

1.3 Thesis Outline

This thesis is mainly divided in two parts. The first part is dealing with ANCOVA analysis for complete data and in the second part, the ideas of Akritas et. all (2000) are extended to factorial designs with repeated measurements when some of the data is missing.

Factorial designs with paired data subject to missingness are presented in Chapter 2. Here the MCR assumption introduced by Akritas et. al (2002) is generalized for cases were there are more than two repeated measurements per subject. This chapter contains asymptotic results for testing main and interaction effects as well as simulation studies and real data analysis using the model proposed. Generalization of this theory to factorial designs with more than two repeated measurements, related simulation studies and data analysis are presented in Chapter 3.

Chapter 4 describes the nonparametric ANCOVA model. This includes defining the model, construction of the test statistics, main results regarding the asymptotic distribution of the test statistics under the null and local alternative hypotheses and their proofs. Also, simulation studies conducted to evaluate the performance of this analysis are presented in this chapter.

Combining the two main methodologies presented in this thesis, mixed effects models with missing data and ANCOVA, there is a potential to developed an ANCOVA model with depended data subject to missingness. In order to develop the later analysis there are are several intermediate models to be investigated. Chapter 5 is the conclusion of this thesis summarizing these new methodologies and discussing future research directions.

Finally, note that in order to achieve a unified presentation of the models and test procedures for all ordinal data, in the rest of this thesis all distribution functions (also conditional ones) are the average of their left- and right-continuous version, $F(x) = \frac{1}{2}[F^+(x) + F^-(x)]$.

Chapter 2

Factorial Designs with MCR Paired Data

2.1 Introduction

In social sciences and medical research, data are often longitudinal and as such they typically contain missing values. Methods for missing data are most often based on parametric models and assumptions. Recently developed nonparametric models and hypotheses for factorial designs with complete data have made it possible to extend the popular Mann-Whitney-Wilcoxon rank sum test and the Kruskal-Wallis test to multifactor designs; see Akritas and Arnold (1994), Akritas et.al. (1997) and Akritas and Brunner (1997). These multifactor extensions share the good power properties and the protection against outliers that the one-factor rank statistics are known for.

The use of rank test statistics with data containing missing values is much less common than in the complete-data case. Their use is often limited to complete-case analysis, which is both inefficient and overly restrictive in its assumptions. In this chapter we consider in detail the case where subjects are observed at two follow-up times, and construct rank statistics for factor effects and interactions under a new missingness assumption, called missingness conditionally at random (MCR), which is neither stronger nor weaker than the classical missingness at random (MAR) assumption Rubin (1976).

The particular application we will consider (see Section 2.5) concerns the study of incarcerated boys in Michigan juvenile correction institutions by Gold and Osgood (1992). The boys were 12-18 years old when entering the institutions. 91% of them participated in a first interview within 10 days of their arrival at the institution. 97% completed an interview four months after their arrival, 91% completed another shortly before departure, and 75% completed another six months later. To address the question of whether apparent prosocial change during incarceration maintains after reentering the community, we analyzed measures of delinquent values obtained when boys left the institutions and six months later. The boys are categorized according to race, the number of prior placements and length of the most recent stay at an institution. All variables are dichotomized. The data set has $N = 336$ observations, out of which both measurements are missing for 24 subjects and the first and second measurements are missing for 6 and 59 subjects, respectively.

Let (Y^0_{i1k}, Y^0_{i2k}), $k = 1, \dots, N_i$, be independent repetitions of the pair of variables of interest $\mathbf{Y}^0_i = (Y^0_{i1}, Y^0_{i2})$, where the possibly vector index i denotes the factor-level combinations, excluding the factor time which is denoted by the second subscript. We will work with a completely nonparametric marginal model that only specifies

$$Y^0_{ijk} \sim F_{ij}, \text{ for all } i, j \text{ and } k. \tag{2.1.1}$$

Thus the nonparametric model assumes only that the marginal distribution of an observation in group i at time point j depends only on i and j.

In this nonparametric context, main effects and interactions are defined through a decomposition of the distribution functions introduced in Akritas and

Arnold (1994). For univariate i, this decomposition is

$$F_{ij}(y) = M(y) + A_i(y) + B_j(y) + C_{ij}(y), \tag{2.1.2}$$

and the decomposition is unique under a set of constraints such as $\sum_i A_i(y) = 0$, $\sum_i B_j(y) = 0$, $\sum_i C_{ij}(y) = 0$, and $\sum_i C_{ij}(y) = 0$ for all i, j and x. The functions A_i, B_j and C_{ij} are, respectively, the nonparametric main effects for factor, main effect for time, and factor-time interaction. The nonparametric hypotheses of no main effects and no interactions simply specify that the corresponding nonparametric effects are zero.

Assume we observe

$$\mathbf{Z}_{ik} = (Y_{i1k}, \Delta_{i1k}, Y_{i2k}, \Delta_{i2k}), \quad k = 1, \ldots, N_i, \tag{2.1.3}$$

where $\Delta_{i1k} = 1$ if Y^0_{i1k} is observed, in which case $Y_{i1k} = Y^0_{i1k}$, and $\Delta_{i1k} = 0$ if Y^0_{i1k} is missing, in which case Y_{i1k} is set to some arbitrary value. The variables Δ_{i2k} and Y_{i2k} are defined analogously. Also, for each factor-level combination i let

$$n_i = \sum_{k=1}^{N_i} \{\Delta_{i1k} + \Delta_{i2k}(1 - \Delta_{i1k})\}, \quad n_{ij} = \sum_{k=1}^{N_i} \Delta_{ijk}, \; i = 1, 2, \quad n_{ci} = \sum_{k=1}^{N_i} \Delta_{i1k}\Delta_{i2k}$$

denote the number of observations with at least one non-missing observation, the number of non-missing observations for each of the two variables, and the number of complete pairs, respectively. Finally, we set $N = \sum_i N_i$ and $n_c = \sum_i n_{ci}$.

The key ingredient for testing hypotheses regarding the nonparametric effects in (2.1.2) is estimation of F_{ij}. In the case of complete observations, this is accomplished by the usual empirical distribution function. If missingness is completely at random, i.e. $(\Delta_{i1k}, \Delta_{i2k})$ is independent of (Y^0_{i1k}, Y^0_{i2k}), this estimation is also easily accomplished by the usual empirical distribution function of the available observations. The real challenge with missing data is to estimate

F_{ij} when missingness is not completely at random. For example, the empirical distribution function of the observed Y^0_{i1k}'s,

$$\hat{F}_{i1}\left(y|\Delta_{i1}=1\right) \;=\; n_{i1}^{-1}\sum_{k=1}^{N_i}\Delta_{i1k}c(Y_{i1k},y), \tag{2.1.4}$$

where

$$c(y,x)=\frac{1}{2}\left\{I(y\le x)+I(y<x)\right\}, \tag{2.1.5}$$

and I denotes an indicator function, is always an unbiased estimator of

$$F_{i1}\left(y|\Delta_{i1}=1\right) \;=\; \frac{1}{2}\{P(Y_{i1}\le y|\Delta_{i1}=1)+P(Y_{i1}<y|\Delta_{i1}=1)\}.$$

If the missingness is completely at random then $F_{i1}\left(y|\Delta_{i1}=1\right) = F_{i1}(y)$ and thus the empirical distribution in (2.1.4) also estimates F_{i1}. This is the basic reason why the aforementioned complete-case analysis works under the assumption that missingness is completely at random. Even then, however, this estimator uses only part of the information in the data. In addition, the assumption of missing completely at random is more restrictive than necessary. The proposed nonparametric approach will remedy both of these shortcomings.

It turns out that the MAR assumption is not suitable for nonparametric analysis. Instead, we will use an alternative missingness assumption, MCR. It will be seen that under MCR, consistent estimation of F_{i1} is accomplished by a type of nonparametric imputation that integrates conditional distributions instead of imputing values from them.

The MCR assumption and the method for estimating F_{ij} are described in Section 2.2. In Sections 2.3, a general test statistic for nonparametric hypotheses involving an arbitrary number of factors is presented and its asymptotic distribution obtained. Results from simulations studies are presented in Section 2.4.

Analyses of the sociological data set are given in Section 2.5, and finally the proofs are presented in Section 2.6.

2.2 MCR and Nonparametric Imputation

In this section we will introduce the MCR assumption and give an estimator of F_{ij} using nonparametric imputation. Let $\mathbf{Y}_i^0 = (Y_{i1}^0, Y_{i2}^0)$ be the pair of variables of interest, as before. The MCR assumption specifies that, for each factor level combination i,

$$P(\Delta_{i1} = 0, \Delta_{i2} = 0 | \mathbf{Y}_i^0) = P(\Delta_{i1} = 0, \Delta_{i2} = 0) \tag{2.2.1}$$

$$P(\Delta_{i1} = 1 | \Delta_{i2} = 1, \mathbf{Y}_i^0) = P(\Delta_{i1} = 1 | \Delta_{i2} = 1, Y_{i2}^0) \tag{2.2.2}$$

$$P(\Delta_{i2} = 1 | \Delta_{i1} = 1, \mathbf{Y}_i^0) = P(\Delta_{i2} = 1 | \Delta_{i1} = 1, Y_{i1}^0). \tag{2.2.3}$$

Remark 2.2.1. *Let $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$ denote those values of the variables that are observed and missing in a sample, and let $\boldsymbol{\Delta}$ denote the vector of missingness indicators $(\Delta_{i1k}, \Delta_{i2k})$ for the whole sample. Then the classical MAR states that*

$$P(\boldsymbol{\Delta} = \boldsymbol{\delta} | \mathbf{Y}_{obs} = \mathbf{Y}_{obs}, \mathbf{Y}_{mis}) = P(\boldsymbol{\Delta} = \boldsymbol{\delta} | \mathbf{Y}_{obs} = \mathbf{Y}_{obs}), \tag{2.2.4}$$

where $\boldsymbol{\delta}$ and $\mathbf{Y}_{obs}$ are the realizations of $\boldsymbol{\Delta}$ and $\mathbf{Y}_{obs}$ in the sample actually observed. This is the weakest general condition under which the missingness mechanism can be ignored in a likelihood-based analysis. Because (2.2.4) refers only to the exact values of $\boldsymbol{\delta}$ and $\mathbf{Y}_{obs}$ in the data at hand, it allows missingness mechanisms that are highly specific to a given set of observations. Often it is more natural to think of MAR in the following more restrictive form. Let $\boldsymbol{\Delta}_i = (\Delta_{i1}, \Delta_{i2})$, and let $\mathbf{Y}_{i,obs}$ and $\mathbf{Y}_{i,mis}$ denote the observed and missing components of $\mathbf{Y}_i^0$ for a given observation. Then the more restrictive version of MAR states that

$$P(\boldsymbol{\Delta}_i | \mathbf{Y}_i^0) = P(\boldsymbol{\Delta}_i | \mathbf{Y}_{i,obs}) \tag{2.2.5}$$

for all observations, with $\mathbf{\Delta}_i$ *for different observations assumed conditionally independent. Comparing (2.2.1)–(2.2.3) with (2.2.5), it is seen that the MCR assumption mainly differs from the classical one in that a component of the* $\mathbf{\Delta}_i$ *vector is part of the conditioning information. Thus (2.2.5) states that conditionally on* $\mathbf{Y}_{i,obs}$, $\mathbf{\Delta}_i$ *is independent from* $\mathbf{Y}_{i,mis}$, *whereas (2.2.3) states that* Δ_{i2} *is independent from* Y_{i2}^0 *given the value of* Y_{i1}^0 *and the fact that* $\Delta_{i1} = 1$. *Finally, (2.2.1) simply says that the pairs with both variables missing do not contain information about the joint distribution of* $\mathbf{Y}_i^0$ *and can thus be ignored. This assumption is also implicit in both forms of the MAR assumption.*

The MCR assumption is neither stronger nor weaker than the full form of MAR. Thus (2.2.4) allows for missingness mechanisms which do not satisfy (2.2.1)–(2.2.3), but MCR also allows models which violate (2.2.4). However, in some cases the assumptions coincide. In the simulations of Section 2.4 we consider two missingness models which satisfy both MCR and the stricter MAR (2.2.5). The first of these is the missing completely at random (MCAR) case where $\mathbf{\Delta}_i$ is independent of $\mathbf{Y}_i$, and the second is the situation where only Y_{i2}^0 may be missing, with the probability of missingness depending on Y_{i1}^0.

The nonparametric estimation of F_{i1} is based on the decomposition

$$
\begin{aligned}
F_{i1}(y) &= F_{i1}(y|\Delta_{i1} = 1 \text{ or } \Delta_{i2} = 1) \\
&= F_{i1}(y|\Delta_{i1} = 1)\, P(\Delta_{i1} = 1|\Delta_{i1} = 1 \text{ or } \Delta_{i2} = 1) \\
&+ F_{i1}(y|\Delta_{i2} = 1, \Delta_{i1} = 0)\, P(\Delta_{i2} = 1, \Delta_{i1} = 0|\Delta_{i1} = 1 \text{ or } \Delta_{i2} = 1) \quad (2.2.6)
\end{aligned}
$$

where $F_{i1}(y|\Delta_{i2} = 1, \Delta_{i1} = 0)$ denotes the conditional distribution of Y_{i1}^0 given $\Delta_{i2} = 1, \Delta_{i1} = 0$. Note that the first equality is due to (2.2.1) which implies that the event $\{\Delta_{i1} = 0, \Delta_{i2} = 0\}$, and thus also its complement

$\{\Delta_{i1} = 1 \text{ or } \Delta_{i2} = 1\}$, is independent from (Y_{i1}^0, Y_{i2}^0). Since $F_{i1}(y|\Delta_{i1} = 1)$ is estimated by (2.1.4), and the probabilities $P(\Delta_{i1} = 1|\Delta_{i1} = 1 \text{ or } \Delta_{i2} = 1)$ and $P(\Delta_{i2} = 1, \Delta_{i1} = 0|\Delta_{i1} = 1 \text{ or } \Delta_{i2} = 1)$ are readily estimated by n_{i1}/n_i and $(n_{i2} - n_{ci})/n_i$, respectively, it follows that estimation of F_{i1} through (2.2.6) can be achieved if we can estimate $F_{i1}(y|\Delta_{i2} = 1, \Delta_{i1} = 0)$. Estimation of the distribution of Y_{i1}^0 conditional on the set where only Y_{i2}^0 is observed, can in principle be accomplished by utilizing the information that Y_{i2}^0 carries about Y_{i1}^0. Utilization of such information nonparametrically is called nonparametric imputation. Our version of nonparametric imputation is based on the relations

$$F_{i1}(y|\Delta_{i2} = 1, \Delta_{i1} = 0) = \int F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 0)dF_{i2}(x|\Delta_{i2} = 1, \Delta_{i1} = 0), \quad (2.2.7)$$

where

$$F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 0) = F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 1). \quad (2.2.8)$$

Here (2.2.8) follows from (2.2.2) since, as already mentioned in Remark 2.2.1, an equivalent restatement of (2.2.2) is that Δ_{i1} is conditionally independent from Y_{i1}^0, given Y_{i2}^0 and $\Delta_{i2} = 1$. Thus, (2.2.2) is the weakest assumption under which (2.2.8) holds. Clearly, (2.2.8) is critical to our nonparametric imputation since it reduces the problem to estimating $F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 0)$, to that of estimating $F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 1)$. The latter is feasible because it calls for estimating the distribution of Y_{i1}^0 in a set where Y_{i1}^0 is observed. Note that the MAR assumption is not tailored for nonparametric imputation, because (2.2.4) does not in general imply (2.2.2).

Estimation of $F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 1)$, and hence of $F_{i1}(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 0)$, can be readily done using smoothing techniques:

$$\widehat{F}_{i1}\left(y|Y_{i2}^0 = x, \Delta_{i2} = 1, \Delta_{i1} = 0\right) = \sum_k w_{i2k}(x)c(Y_{i1k}, y), \tag{2.2.9}$$

where the function c is defined in (2.1.5) and

$$w_{i2k}(x) = \frac{K_{b_{i2}}(x - Y_{i2k})\Delta_{i2k}\Delta_{i1k}}{\sum_l K_{b_{i2}}(x - Y_{i2l})\Delta_{i2l}\Delta_{i1l}}, \tag{2.2.10}$$

where $K_b(x) = b^{-1}K(x/b)$, K is a symmetric kernel function and b is the bandwidth (see, e.g., Wand and Jones (1995) and Akritas et. al. (2000)). Using this, relation (2.2.7) implies that

$$\begin{aligned} &\widehat{F}_{i1}\left(y|\Delta_{i2} = 1, \Delta_{i1} = 0\right) \\ &= \frac{1}{n_{i2} - n_{ci}} \sum_{k=1}^{N_i} \Delta_{i2k}(1 - \Delta_{i1k})\widehat{F}_{i1}(y|Y_{i2}^0 = Y_{i2k}, \Delta_{i2} = 1, \Delta_{i1} = 0). \end{aligned} \tag{2.2.11}$$

Finally, relation (2.2.6) implies that

$$\begin{aligned} \widehat{F}_{i1}(y) = \frac{1}{n_i} \sum_{k=1}^{N_i} &\{\Delta_{i1k} c\left(Y_{i1k}, y\right) \\ &+ \Delta_{i2k}(1 - \Delta_{i1k})\widehat{F}_{i1}\left(y|Y_{i2}^0 = Y_{i2k}, \Delta_{i2} = 1, \Delta_{i1} = 0\right)\Big\} \end{aligned} \tag{2.2.12}$$

is a consistent estimator of F_{i1}. A consistent estimator $\widehat{F}_{i2}$ of F_{i2} is defined similarly.

Remark 2.2.2. *(a) This version of nonparametric imputation is slightly different than the standard one, in that we do not impute values of Y_{i1k}^0 from the conditional distribution in (2.2.9). Instead, we average (integrate) these conditional distributions over all observed values of Y_{i2k}^0 with $\Delta_{i2k} = 1, \Delta_{i1k} = 0$.*
(b) If the missingness is completely at random then $F_{i1}(y|\Delta_{i2} = 1, \Delta_{i1} = 0) = F_{i1}(y)$, and thus (2.2.11) also estimates $F_{i1}(y)$. Of course, even under the assumption that the missingness if completely at random, one would use (2.2.12),

which combines the estimators in (2.1.4) and in (2.2.11), in order to use all the information contained in the sample. Simulations performed under missingness completely at random in Section 2.4 indicate that tests based on (2.2.12) are indeed more powerful that the complete-case analysis which uses only (2.1.4).

2.3 Test procedures

Consider the data notation introduced in Section 2.1, and the fully nonparametric model and hypotheses described in (2.1.1) and (2.1.2), and let $\mathbf{F}$ denote the vector of the marginal cumulative distribution functions F_{ij}. A unified presentation of test procedures for any such nonparametric hypothesis is made possible from the observation that they can all be expressed in the form

$$H_0: \ \mathbf{CF} = \mathbf{0}, \tag{2.3.1}$$

where $\mathbf{C}$ is a contrast matrix. For example, in the case of univariate index i, $\mathbf{F} = (F_{11}, F_{12}, \ldots, F_{I1}, F_{I2})'$, where I denotes the total number of factor levels. In this case, the hypotheses of interest are given in terms of the effects in the decomposition (2.1.2):

$$H_0(F): \ A_i = 0; \quad H_0(T): \ B_j = 0; \quad H_0(TF): \ C_{ij} = 0$$
$$H_0(T|F): \ A_i + C_{ij} = 0; \quad H_0(F|T): \ B_j + C_{ij} = 0$$

for all i and j. Letting $\mathbf{M}_d = (\mathbf{1}_{d-1}|\mathbf{I}_{d-1})$ where $\mathbf{1}_{d-1}$ denotes the $(d-1)\times 1$ vector of ones, and $\mathbf{I}_{d-1}$ is the $d-1$ dimensional identity matrix, the contrast matrices for the four hypotheses above are given by $\mathbf{C}(F) = \mathbf{M}_I \otimes (\frac{1}{2}\mathbf{1}')$, $\mathbf{C}(T) = (I^{-1}\mathbf{1}_I') \otimes \mathbf{M}_2$, $\mathbf{C}(TF) = \mathbf{M}_I \otimes \mathbf{M}_2$, $\mathbf{C}(T|F) = \mathbf{I}_I \otimes \mathbf{M}_2$ and $\mathbf{C}(F|T) = \mathbf{M}_I \otimes \mathbf{I}_2$ respectively.

The test procedures we will consider for these hypotheses are extensions of the rank procedures in Akritas and Arnold (1994) and Akritas and Brunner

(1997) with complete data. For $i = 1, \ldots, I$, let $\hat{F}_{i1}$ be as defined in (2.2.12) and $\hat{F}_{i2}$ be defined similarly, let $n_{\cdot\cdot} = \sum_{i=1}^{I} \sum_{j=1}^{2} n_{ij}$, and set

$$\widehat{H}(y) = \frac{1}{n_{\cdot\cdot}} \sum_{i=1}^{I} \sum_{j=1}^{2} \sum_{k=1}^{N_i} \Delta_{ijk} c(Y_{ijk}, y) \tag{2.3.2}$$

for the empirical distribution function obtained from all data. The proposed test statistic for the hypothesis H_0 given in (2.3.1) is based on $T_{\mathbf{C}} = \mathbf{C} \int \widehat{H}(y) d\widehat{\mathbf{F}}(y)$. In this section we will derive the asymptotic distribution of $T_{\mathbf{C}}$. As a first step we note that under the null hypothesis (2.3.1), this is equal to

$$T_{\mathbf{C}} = \mathbf{C} \int \widehat{H}(y) d\left\{\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right\}. \tag{2.3.3}$$

In view of the form (2.3.3) it follows that a unified theory for test statistics for the nonparametric hypotheses in factorial designs with arbitrarily many factors is possible from the asymptotic distribution of the vector

$$\int \widehat{H}(y) d\left\{\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right\}. \tag{2.3.4}$$

Proposition 2.3.1. *Let $\widehat{H}$ be defined by (3.3.1) and set $H = E(\widehat{H})$. Also let $\widehat{F}_{ij}$ be the estimator defined in (2.2.12). Then, for each $i = 1, \ldots, I$ and $j = 1, 2$, and under Assumptions A1-A3 stated in Section 2.6,*

$$N^{1/2} \int \left(\widehat{H} - H\right) d\left(\widehat{F}_{ij} - F_{ij}\right) \to 0,$$

in probability, as $N \to \infty$.

The proof is given in Section 2.6.

Proposition 2.3.1 implies that $\int \{\widehat{H}(y) - H(y)\} d\left\{\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right\} = o_p(N^{-1/2})$, and thus the asymptotic distribution of (2.3.4) follows from that of

$$\int H(y) d\left\{\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right\}. \tag{2.3.5}$$

Note that the elements of $\widehat{\mathbf{F}}$ consist of $(\widehat{F}_{i1}, \widehat{F}_{i2})$, $i = 1, \ldots, I$, and these pairs are independent. Thus, the asymptotic distribution of (2.3.5) is implied by that of

$$\left(\int H(y)d\left\{\widehat{F}_{i1}(y) - F_{i1}(y)\right\}, \int H(y)d\left\{\widehat{F}_{i2}(y) - F_{i2}(y)\right\}\right),$$

which is given in Corollary 2.3.3. First we need the following representation.

Theorem 2.3.2. *Let $\widehat{F}_{ij}$ be the estimator defined in (2.2.12). Then, for each $i = 1, \ldots, I$ and $j = 1, 2$, and under Assumptions A1-A3 stated in Section 2.6,*

$$\begin{aligned}
&\int H(y)d\left\{\widehat{F}_{i1}(y) - F_{i1}(y)\right\} \\
= \; & \frac{1}{n_i}\sum_{k=1}^{N_i}\{\Delta_{i1k} + \Delta_{i2k}(1-\Delta_{i1k})\}\left[H(Y_{i1k}^0) - E\{H(Y_{i1k}^0)\}\right] \\
&+ \frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left[E_1^*\{H(Y_{i1}^0)|Y_{i2k}\} - H(Y_{i1k}^0)\right] \\
&+ \frac{2}{n_{ci}}\sum_{k=1}^{N_i} H_{1,1}(\mathbf{Z}_{ik}) + o_p\left(N^{-1/2}\right),
\end{aligned}$$

as $N \to \infty$, where $\mathbf{Z}_{ik}$ is defined in (2.1.3)*,*

$$\begin{aligned}
E_1^*\{H(Y_{i1}^0)|Y_{i2k}\} &= E\{H(Y_{i1}^0)|Y_{i2}^0 = Y_{i2k}, \Delta_{i1} = 0, \Delta_{i2} = 1\}, \\
H_{1,1}(\mathbf{Z}_{ik}) &= E\left\{H_1\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right)|\mathbf{Z}_{ik}\right\} \text{ with} \\
H_1\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right) &= \frac{1}{2}\left\{\widetilde{H}_1\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right) + \widetilde{H}_1\left(\mathbf{Z}_{il}, \mathbf{Z}_{ik}\right)\right\}, \text{ where}
\end{aligned}$$

$$\begin{aligned}
\widetilde{H}_1\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right) &= \Delta_{i2k}(1-\Delta_{i1k})\frac{K_{b_{i2}}(Y_{i2k} - Y_{i2l})\Delta_{i1l}\Delta_{i2l}}{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2} = 1)} \\
&\quad \times \left[H(Y_{i1l}^0) - E_1^*\{H(Y_{i1}^0)|Y_{i2l}\}\right],
\end{aligned}$$

and $f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2} = 1)$ is the conditional density of Y_{i2} given $\Delta_{i1}\Delta_{i2} = 1$ evaluated at Y_{i2k}.

The proof is given in Section 2.6.

Define now

$$
\begin{aligned}
h_{i1}(\mathbf{Z}_{ik}) &= \{\Delta_{i1k} + \Delta_{i2k}(1-\Delta_{i1k})\}\left[H(Y^0_{i1k}) - E\{H(Y^0_{i1k})\}\right] \qquad (2.3.6)\\
&\quad +\Delta_{i2k}(1-\Delta_{i1k})\left[E^*_1\{H(Y^0_{i1})|Y_{i2k}\} - H(Y^0_{i1k})\right] + 2\frac{n_i}{n_{ci}}H_{1,1}(\mathbf{Z}_{ik}),\\
h_{i2}(\mathbf{Z}_{ik}) &= \{\Delta_{i2k} + \Delta_{i1k}(1-\Delta_{i2k})\}\left[H(Y^0_{i2k}) - E\{H(Y^0_{i2k})\}\right] \qquad (2.3.7)\\
&\quad +\Delta_{i1k}(1-\Delta_{i2k})\left[E^*_2\{H(Y^0_{i2})|Y_{i1k}\} - H(Y^0_{i2k})\right] + 2\frac{n_i}{n_{ci}}H_{2,1}(\mathbf{Z}_{ik}),
\end{aligned}
$$

where $E^*_2\{H(Y^0_{i2})|Y_{i1k}\} = E\{H(Y^0_{i2})|Y^0_{i1} = Y_{i1k}, \Delta_{i1} = 1, \Delta_{i2} = 0\}$, and $H_{2,1}(\mathbf{Z}_{ik})$ is defined in a way analogous to the definition of $H_{1,1}(\mathbf{Z}_{ik})$ given in Theorem 2.3.2; i.e. $H_{2,1}(\mathbf{Z}_{ik}) = E\left\{H_2\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right)|\mathbf{Z}_{ik}\right\}$ with $H_2\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right) = \frac{1}{2}\left\{\widetilde{H}_2\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right) + \widetilde{H}_2\left(\mathbf{Z}_{il}, \mathbf{Z}_{ik}\right)\right\}$, where

$$
\begin{aligned}
\widetilde{H}_2\left(\mathbf{Z}_{ik}, \mathbf{Z}_{il}\right) &= \Delta_{i1k}(1-\Delta_{i2k})\frac{K_{b_{i1}}(Y_{i1k} - Y_{i1l})\Delta_{i2l}\Delta_{i1l}}{f_{i1}(Y_{i1k}|\Delta_{i1}\Delta_{i2} = 1)}\\
&\quad \times\left[H(Y^0_{i2l}) - E^*_2\{H(Y^0_{i2})|Y_{i1l}\}\right],
\end{aligned}
$$

and $f_{i1}(Y_{i1k}|\Delta_{i1}\Delta_{i2} = 1)$ is the conditional density of Y_{i1} given $\Delta_{i1}\Delta_{i2} = 1$ evaluated at Y_{i1k}. With this notation we have

Corollary 2.3.3. *Let the assumptions of Theorem 2.3.2 hold, and let $h_{ij}(\mathbf{Z}_{ik})$, $j = 1, 2$, be defined in* (2.3.6)*.*

1. *Let $\mathbf{\Sigma}_i$ be the 2×2 covariance matrix with elements $\sigma_{11i} = \mathrm{var}\{h_{i1}(\mathbf{Z}_{ik})\}$, $\sigma_{22i} = \mathrm{var}\{h_{i2}(\mathbf{Z}_{ik})\}$, $\sigma_{12i} = \mathrm{cov}\{h_{i1}(\mathbf{Z}_{ik}), h_{i2}(\mathbf{Z}_{ik})\}$. Then,*
$$
n_i^{1/2}\left(\int H(y)d\left\{\widehat{F}_{i1}(y) - F_{i1}(y)\right\}, \int H(y)d\left\{\widehat{F}_{i2}(y) - F_{i2}(y)\right\}\right) \to N(\mathbf{0}, \mathbf{\Sigma}_i).
$$

2. *Let $(N/n_i) \to \lambda_i$, as $N \to \infty$, $i = 1, \ldots, I$, and let $\mathbf{V}$ be the block diagonal covariance matrix with the i-th block being $\lambda_i\mathbf{\Sigma}_i$. Then, as $N \to \infty$,*
$$
N^{1/2}\int \widehat{H}(y)d\left\{\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right\} \to N(\mathbf{0}, \mathbf{V}),
$$

in distribution.

To obtain a consistent estimator of $\mathbf{V}$, we first obtain consistent estimators of $h_{ij}(\mathbf{Z}_{ik}),\ j = 1, 2$, and then we estimate $\mathbf{\Sigma}_i$ by the sample covariance matrix of $(h_{i1}(\mathbf{Z}_{ik}), h_{i2}(\mathbf{Z}_{ik}))$, $k = 1, \ldots, n_i$. Consistent estimation of

$$\begin{aligned} h_{i1}(\mathbf{Z}_{ik}) &= \Delta_{i1k} H(Y_{i1k}) - \{\Delta_{i1k} + \Delta_{i2k}(1-\Delta_{i1k})\} E\{H(Y_{i1}^0)\} \\ &\quad + \Delta_{i2k}(1-\Delta_{i1k}) E_1^*\{H(Y_{i1}^0)|Y_{i2k}\} + 2\frac{n_i}{n_{ci}} H_{1,1}(\mathbf{Z}_{ik}) \end{aligned}$$

can be done by consistently estimating each of its terms. The first term can be consistently estimated by $\Delta_{i1k}\widehat{H}(Y_{i1k})$, where $\widehat{H}(y)$ is defined in (3.3.1). The expectation in the second term can be consistently estimated by

$$\begin{aligned} \widehat{E}\{H(Y_{i1}^0)\} &= \int \widehat{H}(y) d\widehat{F}_{i1}(y) \\ &= \frac{1}{n_i}\sum_{l=1}^{N_i}\left\{\Delta_{i1l}\widehat{H}(Y_{i1l}) + \Delta_{i2l}(1-\Delta_{i1l})\sum_{m=1}^{N_i} w_{i2m}(Y_{i2l})\widehat{H}(Y_{i1m})\right\}, \end{aligned}$$

the expectation in the third term can be consistently estimated by

$$\begin{aligned} \widehat{E}_1^*\{H(Y_{i1}^0)|Y_{i2k}\} &= \int \widehat{H}(y) d\widehat{F}_{i1}(y|Y_{i2}^0 = Y_{i2k}, \Delta_{i1} = 0, \Delta_{i2} = 1) \\ &= \sum_{l=1}^{N_i} w_{i2l}(Y_{i2k})\widehat{H}(Y_{i1l}), \end{aligned}$$

and finally, a consistent estimator of $H_{1,1}(\mathbf{Z}_{ik})$ is

$$\widehat{H}_{1,1}(\mathbf{Z}_{ik}) = \frac{1}{2n_i}\sum_{l=1}^{N_i}\left\{\widehat{\widehat{H}}_1(\mathbf{Z}_{ik}, \mathbf{Z}_{il}) + \widehat{\widehat{H}}_1(\mathbf{Z}_{il}, \mathbf{Z}_{ik})\right\}$$

where

$$\begin{aligned} \widehat{\widehat{H}}_1(\mathbf{Z}_{ik}, \mathbf{Z}_{il}) &= \Delta_{i2k}(1-\Delta_{i1k})\frac{K_{b_{i2}}(Y_{i2k} - Y_{i2l})\Delta_{i1l}\Delta_{i2l}}{\widehat{f}_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2} = 1)} \\ &\quad \times\left[\widehat{H}(Y_{i1l}) - \widehat{E}_1^*\{H(Y_{i1}^0)|Y_{i2l}\}\right], \end{aligned}$$

with $\hat{f}_{i2}(y|\Delta_{i1}\Delta_{i2} = 1) = \frac{1}{n_c}\sum_{m=1}^{N_i} \Delta_{i1m}\Delta_{i2m}K_{b_{i2}}(y - Y_{i2m})$. Similarly, we can obtain a consistent estimator for $h_{i2}(\mathbf{Z}_{ik})$.

Let $\widehat{\boldsymbol{\Sigma}}_i$, denote the aforementioned estimator of $\boldsymbol{\Sigma}_i$, and set $\widehat{\mathbf{V}}$ for the block diagonal matrix with diagonal elements $(N/n_i)\widehat{\boldsymbol{\Sigma}}_i$.

Corollary 2.3.4. *Consider the notation and assumptions of Corollary 2.3.3, and let* $\mathbf{C}$ *be any* $r \times I$ *full row rank contrast matrix. Then as* $N \to \infty$,

$$N \left\{ \int \widehat{H}(y) d\widehat{\mathbf{F}}'(y)\mathbf{C}' \right\} (\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}')^{-1} \left\{ \mathbf{C} \int \widehat{H}(y) d\widehat{\mathbf{F}}(y) \right\} \to \chi_r^2,$$

in distribution.

2.4 Simulations

When the data are MAR, the likelihood is obtained by collapsing the joint distribution of the observed, Y_{obs}, and the missing parts of the data over the missing part. This is called *the likelihood ignoring the missing data mechanism* (Little and Rubin, 1987) or *the observed data likelihood* (Schafer, 1997). With our assumption that the pairs (Y_{i1k}, Y_{i2k}) follow a bivariate normal distribution with $E(Y_{ijk}) = \mu_{ij}$, $Var(Y_{ijk}) = \sigma_{ij}^2$, and $Cov(Y_{i1k}, Y_{i2k}) = \sigma_{i,12}$, for $i = 1, \ldots, I,\ j = 1, 2,\ k = 1, \ldots, N_i$, the logarithm of the observed data likelihood is

$$\begin{aligned} l(\boldsymbol{\Theta}\ ; Y_{obs}) \ = \ & \sum_{i=1}^{I}\sum_{k=1}^{N_i} [\Delta_{i1k}\Delta_{i2k} \log f(Y_{i1k}, Y_{i2k}; \theta_{i1}, \theta_{i2}, \sigma_{i,12}) \\ & +\Delta_{i1k}(1 - \Delta_{i2k}) \log f(Y_{i1k}; \theta_{i1}) \\ & + \Delta_{i2k}(1 - \Delta_{i1k}) \log f(Y_{i2k}; \theta_{i2})] \,, \end{aligned} \tag{2.4.1}$$

where we set $\theta_{ij} = (\mu_{ij}, \sigma_{ij}^2)$, $\boldsymbol{\Theta} = (\theta_{11}, \ldots, \theta_{I2}, \sigma_{1,12}, \ldots, \sigma_{I,12})$ and, with self-understood notation, f denotes both the bivariate and univariate normal density. We will apply the likelihood ratio test for testing hypotheses of the form H_0^{μ} :

$\mathbf{C}\boldsymbol{\mu} = 0$, where $\mathbf{C}$ is a contrast matrix and $\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{I2})'$. Note that such hypotheses are implied by their nonparametric counterparts (2.3.1), and that the common hypotheses of no main group effect, no main time effect, and no interaction, are of this form.

In our simulations we consider a one-factor design with two levels. Thus we have $I = 2$ groups crossed with the two time points. The likelihood ratio statistic

$$2\left[l(\widehat{\boldsymbol{\Theta}}_a\ ; Y_{obs}) - l(\widehat{\boldsymbol{\Theta}}_0\ ; Y_{obs})\right],$$

where $\widehat{\boldsymbol{\Theta}}_a$ and $\widehat{\boldsymbol{\Theta}}_0$ are the maximizers of (2.4.1) over the parameter space defined by the alternative and the null hypotheses respectively, was constructed under the additional hypothesis of homoscedasticity across time points and across groups, that is $\sigma_{ij}^2 = \sigma^2$, and $\sigma_{i,12} = \sigma_{12},\ i = 1, 2,\ j = 1, 2$. Thus under each null hypothesis, the corresponding likelihood ratio statistic follows a χ^2-distribution with one degree of freedom.

In the following simulation study we consider two families of joint distributions to generate the pairs (Y_{i1k}, Y_{i2k}), for $i = 1, 2,\ k = 1, \ldots, N_i$. The first family is a bivariate normal distribution with means $\mu_{i1} = 0,\ \mu_{i2} = \tau[(N_1 + N_2)/2]^{-1/2}$, variances $\sigma_{i1}^2 = \sigma_{i2}^2 = 1$, and correlation ρ_i, for $i = 1, 2$. The second family is bivariate gamma distribution such that the marginal distributions of $Y_{i1k},\ Y_{i2k}$ are Gamma, with location parameters $\mu_{i1} = 1$, and $\mu_{i2} = 1 + \tau[(N_1 + N_2)/2]^{-1/2}$ respectively, scale parameters equal to one, and $Corr(Y_{i1k}, Y_{i2k}) = \rho_i,\ i = 1, 2,\ k = 1, \ldots, N_i$. Note that $\tau = 0$ corresponds to the null hypotheses and values away from zero generate alternatives to the hypothesis of no main time main effect. Under both families of distributions we consider cases where $\rho_1 = \rho_2 = .25$, $\rho_1 = \rho_2 = .75$, and $\rho_1 = .25,\ \rho_2 = .75$.

We consider two models to create missingness patterns satisfying both the MCR and the MAR assumptions. Under the first model, we presume the data is missing completely at random with $P(\Delta_{ijk} = 0) = p_i$, for $i = 1, 2,\ j = 1, 2,$ and $k = 1, \ldots, n_i$. We considered cases where $p_1 = p_2 = 0.1$, $p_1 = p_2 = 0.3$, and $p_1 = 0.1,\ p_2 = 0.3$. Under the second model, only observations from the second time point are missing with a probability depending on the value of the first time point. More specifically, $P(\Delta_{i2j} = 0) = 0.6 - 0.2|Y_{i1k}|$ if $|Y_{i1k}| < 3$ and 0.1 otherwise.

Simulations indicated that the assumption of compact support for the kernel function is not critical for the performance of the procedure, and we chose the normal density kernel for our simulations. In all cases we consider in the simulation study the nominal level is $\alpha = .1$. The R statistical package was used and all the results are based on 1000 runs. Tables 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6 report Type I error rates for the competing procedures and all null hypotheses. For the proposed nonparametric test we used the bandwidths $b_1 = b_2 = .5$, $b_1 = b_2 = 1$ and $b_1 = b_2 = 2$. Power comparisons were done only for main time effect alternatives, and the results are presented in Figures 2.1-2.10.

Cases with same correlation in both of the groups, $\rho_1 = \rho_2$, and same proportion of missingness in both of the groups, $p_1 = p_2$, are summarized in Table 2.1 for normal cases and in Table 2.2 for gamma cases, whereas Table 2.3 summarizes the results for cases with $\rho_1 = .25,\ \rho_2 = .75$ and different combinations of proportions of missingness in each group. For cases generated under the MCR scheme, cases with same correlation in both of the groups, $\rho_1 = \rho_2$ are summarized in Table 2.4 for normal distribution and in Table 2.5 for gamma distribution, and results for cases with $\rho_1 = .25,\ \rho_2 = .75$ are summarized in Table 2.6.

Normal Dist.		20% Incomplete Pairs					60% Incomplete Pairs				
		LR	CP	$NP_{.5}$	NP_1	NP_2	LR	CP	$NP_{.5}$	NP_1	NP_2
ρ	Eff.				$N_1 = 30$		$N_2 = 30$				
0.25	T	0.100	0.111	0.106	0.102	0.101	0.111	0.116	0.123	0.121	0.115
	G	0.100	0.116	0.101	0.099	0.103	0.114	0.123	0.129	0.121	0.124
	TG	0.112	0.109	0.112	0.109	0.107	0.130	0.120	0.149	0.132	0.129
0.75	T	0.108	0.112	0.112	0.108	0.106	0.127	0.117	0.136	0.118	0.119
	G	0.104	0.112	0.102	0.107	0.101	0.100	0.105	0.119	0.123	0.128
	TG	0.110	0.104	0.105	0.105	0.098	0.135	0.122	0.148	0.132	0.128
ρ	Eff.				$N_1 = 50$		$N_2 = 50$				
0.25	T	0.091	0.094	0.090	0.092	0.090	0.108	0.108	0.116	0.104	0.100
	G	0.100	0.105	0.104	0.107	0.105	0.107	0.118	0.118	0.113	0.111
	TG	0.107	0.108	0.105	0.104	0.104	0.115	0.103	0.101	0.106	0.106
0.75	T	0.094	0.091	0.092	0.085	0.091	0.097	0.102	0.108	0.097	0.096
	G	0.100	0.101	0.112	0.111	0.111	0.112	0.119	0.117	0.126	0.123
	TG	0.097	0.113	0.111	0.102	0.101	0.099	0.107	0.100	0.093	0.095
ρ	Eff.				$N_1 = 100$		$N_2 = 100$				
0.25	T	0.104	0.098	0.106	0.107	0.110	0.118	0.113	0.118	0.119	0.117
	G	0.109	0.110	0.107	0.110	0.107	0.112	0.106	0.112	0.115	0.116
	TG	0.097	0.105	0.113	0.114	0.109	0.104	0.109	0.111	0.099	0.099
0.75	T	0.099	0.103	0.101	0.091	0.094	0.119	0.113	0.118	0.104	0.104
	G	0.105	0.098	0.101	0.103	0.097	0.094	0.103	0.091	0.105	0.108
	TG	0.115	0.108	0.116	0.114	0.112	0.118	0.114	0.119	0.111	0.106
ρ	Eff.				$N_1 = 50$		$N_2 = 100$				
0.25	T	0.134	0.114	0.117	0.112	0.110	0.127	0.117	0.116	0.103	0.098
	G	0.129	0.115	0.101	0.102	0.106	0.134	0.125	0.121	0.126	0.124
	TG	0.108	0.104	0.097	0.096	0.098	0.121	0.107	0.117	0.108	0.103
0.75	T	0.132	0.112	0.109	0.100	0.104	0.147	0.116	0.127	0.117	0.113
	G	0.123	0.107	0.108	0.112	0.111	0.130	0.131	0.117	0.121	0.120
	TG	0.127	0.116	0.112	0.106	0.100	0.128	0.105	0.117	0.104	0.097

Table 2.1: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR; same ρ and proportion of missingness in the two groups. Here and all tables T, G, TG denote the hypotheses of no main time effect, no main group effect, no interaction effect, and LR, CP, NP_b denote the likelihood ratio test, the complete pairs test, the proposed nonparametric test with bandwidths $b_1 = b_2 = b$.

Gamma Dist.		20% Incomplete Pairs					60% Incomplete Pairs				
		LR	CP	$NP_{.5}$	NP_1	NP_2	LR	CP	$NP_{.5}$	NP_1	NP_2
ρ	Eff.	$N_1 = 30$					$N_2 = 30$				
0.25	T	0.110	0.091	0.097	0.092	0.092	0.117	0.113	0.109	0.100	0.101
	G	0.103	0.113	0.113	0.114	0.115	0.094	0.113	0.117	0.117	0.116
	TG	0.102	0.114	0.115	0.111	0.111	0.122	0.115	0.132	0.121	0.116
0.75	T	0.110	0.118	0.113	0.109	0.107	0.121	0.117	0.128	0.121	0.122
	G	0.110	0.111	0.110	0.115	0.119	0.113	0.106	0.127	0.126	0.120
	TG	0.093	0.093	0.098	0.103	0.107	0.108	0.109	0.117	0.110	0.104
ρ	Eff.	$N_1 = 50$					$N_2 = 50$				
0.25	T	0.105	0.106	0.102	0.095	0.097	0.092	0.079	0.095	0.087	0.086
	G	0.091	0.098	0.097	0.094	0.100	0.116	0.101	0.112	0.107	0.106
	TG	0.105	0.098	0.109	0.109	0.108	0.115	0.097	0.113	0.107	0.106
0.75	T	0.101	0.101	0.085	0.080	0.082	0.107	0.119	0.112	0.098	0.092
	G	0.102	0.094	0.113	0.109	0.106	0.110	0.107	0.124	0.130	0.127
	TG	0.099	0.108	0.109	0.096	0.094	0.114	0.115	0.125	0.107	0.106
ρ	Eff.	$N_1 = 100$					$N_2 = 100$				
0.25	T	0.095	0.089	0.085	0.086	0.086	0.108	0.095	0.104	0.099	0.096
	G	0.091	0.088	0.098	0.095	0.090	0.089	0.101	0.095	0.093	0.090
	TG	0.122	0.106	0.111	0.108	0.106	0.116	0.100	0.109	0.101	0.096
0.75	T	0.107	0.096	0.097	0.091	0.092	0.128	0.105	0.104	0.101	0.111
	G	0.107	0.113	0.120	0.123	0.121	0.122	0.101	0.116	0.120	0.118
	TG	0.098	0.098	0.099	0.091	0.088	0.104	0.097	0.103	0.103	0.091
ρ	Eff.	$N_1 = 50$					$N_2 = 100$				
0.25	T	0.126	0.107	0.107	0.103	0.103	0.133	0.112	0.112	0.113	0.108
	G	0.192	0.107	0.103	0.103	0.106	0.208	0.103	0.121	0.119	0.120
	TG	0.125	0.113	0.126	0.123	0.122	0.112	0.114	0.115	0.111	0.109
0.75	T	0.116	0.104	0.109	0.100	0.096	0.130	0.114	0.112	0.094	0.093
	G	0.195	0.119	0.122	0.123	0.122	0.195	0.120	0.133	0.136	0.136
	TG	0.117	0.104	0.113	0.104	0.101	0.133	0.113	0.120	0.103	0.100

Table 2.2: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR; same ρ and proportion of missingness in the two groups.

		Normal Distribution					Gamma Distribution				
p_1, p_2	Eff.	LR	CP	$NP_{.5}$	NP_1	NP_2	LR	CP	$NP_{.5}$	NP_1	NP_2
.1, .1	T	0.198	0.104	0.100	0.099	0.105	0.204	0.102	0.104	0.099	0.102
	G	0.106	0.121	0.104	0.111	0.114	0.188	0.105	0.100	0.105	0.108
	TG	0.178	0.101	0.098	0.091	0.096	0.189	0.127	0.121	0.117	0.116
.3, .3	T	0.192	0.111	0.114	0.102	0.100	0.169	0.098	0.112	0.112	0.105
	G	0.118	0.121	0.122	0.127	0.123	0.192	0.103	0.106	0.102	0.104
	TG	0.188	0.105	0.121	0.111	0.105	0.227	0.116	0.130	0.118	0.113
.1, .3	T	0.149	0.110	0.107	0.105	0.101	0.138	0.096	0.103	0.095	0.096
	G	0.111	0.115	0.112	0.116	0.118	0.166	0.102	0.104	0.108	0.102
	TG	0.130	0.095	0.092	0.090	0.093	0.141	0.109	0.116	0.114	0.114

Table 2.3: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR; $N_1 = 50$, $N_2 = 100$, $\rho_1 = .25$ and $\rho_2 = .75$.

Normal	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
Distr.	LR	CP	$NP_{.5}$	NP_1	NP_2	LR	CP	$NP_{.5}$	NP_1	NP_2
Eff.	$N_1 = 30$ $N_2 = 30$									
T	0.096	0.102	0.107	0.101	0.097	0.090	0.097	0.109	0.100	0.100
G	0.115	0.123	0.121	0.121	0.119	0.104	0.126	0.115	0.130	0.125
TG	0.120	0.121	0.137	0.124	0.122	0.119	0.112	0.132	0.120	0.115
Eff.	$N_1 = 50$ $N_2 = 50$									
T	0.091	0.112	0.096	0.094	0.089	0.100	0.112	0.107	0.091	0.092
G	0.105	0.101	0.097	0.101	0.102	0.100	0.108	0.103	0.113	0.112
TG	0.100	0.099	0.116	0.104	0.097	0.098	0.099	0.110	0.106	0.111
Eff.	$N_1 = 100$ $N_2 = 100$									
T	0.099	0.088	0.102	0.099	0.098	0.101	0.101	0.095	0.101	0.107
G	0.105	0.120	0.110	0.113	0.113	0.094	0.104	0.115	0.124	0.119
TG	0.096	0.112	0.100	0.099	0.097	0.095	0.100	0.092	0.106	0.124
Eff.	$N_1 = 50$ $N_2 = 100$									
T	0.119	0.092	0.107	0.101	0.096	0.117	0.107	0.118	0.108	0.111
G	0.138	0.111	0.124	0.120	0.119	0.132	0.102	0.112	0.127	0.120
TG	0.127	0.097	0.115	0.105	0.106	0.121	0.102	0.116	0.113	0.114

Table 2.4: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR; same ρ in the two groups.

Gamma Dist.	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
	LR	CP	$NP_{.5}$	NP_1	NP_2	LR	CP	$NP_{.5}$	NP_1	NP_2
Eff.	$N_1 = 30$ $N_2 = 30$									
T	0.108	0.254	0.111	0.105	0.116	0.111	0.168	0.132	0.173	0.261
G	0.104	0.095	0.116	0.108	0.109	0.113	0.097	0.123	0.127	0.117
TG	0.122	0.101	0.127	0.120	0.115	0.126	0.110	0.137	0.123	0.113
Eff.	$N_1 = 50$ $N_2 = 50$									
T	0.080	0.339	0.121	0.115	0.117	0.113	0.180	0.113	0.209	0.356
G	0.112	0.101	0.106	0.097	0.104	0.088	0.087	0.106	0.108	0.103
TG	0.112	0.097	0.097	0.093	0.093	0.109	0.111	0.121	0.114	0.111
Eff.	$N_1 = 100$ $N_2 = 100$									
T	0.102	0.542	0.099	0.100	0.116	0.096	0.264	0.141	0.315	0.561
G	0.094	0.087	0.093	0.092	0.090	0.100	0.104	0.099	0.098	0.091
TG	0.102	0.101	0.096	0.095	0.092	0.101	0.096	0.102	0.097	0.101
Eff.	$N_1 = 50$ $N_2 = 100$									
T	0.117	0.401	0.110	0.113	0.124	0.131	0.214	0.140	0.241	0.419
G	0.198	0.096	0.109	0.110	0.104	0.174	0.110	0.119	0.121	0.116
TG	0.142	0.116	0.120	0.115	0.111	0.108	0.101	0.103	0.104	0.100

Table 2.5: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR; same ρ in the two groups.

$\rho_1 = .25$	Normal Distribution					Gamma Distribution				
$\rho_2 = .75$	LR	CP	$NP_{.5}$	NP_1	NP_2	LR	CP	$NP_{.5}$	NP_1	NP_2
Eff.				$N_1 = 30$		$N_2 = 30$				
T	0.105	0.104	0.113	0.101	0.098	0.124	0.201	0.129	0.137	0.158
G	0.103	0.120	0.116	0.122	0.119	0.115	0.113	0.102	0.104	0.111
TG	0.133	0.116	0.124	0.123	0.121	0.126	0.124	0.117	0.111	0.116
Eff.				$N_1 = 50$		$N_2 = 50$				
T	0.096	0.118	0.101	0.093	0.090	0.104	0.258	0.110	0.128	0.183
G	0.105	0.097	0.107	0.107	0.105	0.121	0.134	0.116	0.115	0.115
TG	0.112	0.100	0.108	0.103	0.105	0.171	0.147	0.120	0.120	0.129
Eff.				$N_1 = 100$		$N_2 = 100$				
T	0.098	0.104	0.107	0.103	0.102	0.115	0.438	0.131	0.165	0.253
G	0.102	0.116	0.116	0.114	0.121	0.155	0.134	0.128	0.135	0.133
TG	0.111	0.105	0.106	0.105	0.101	0.179	0.191	0.102	0.111	0.139
Eff.				$N_1 = 50$		$N_2 = 100$				
T	0.178	0.101	0.114	0.110	0.102	0.199	0.286	0.115	0.137	0.184
G	0.129	0.109	0.122	0.124	0.115	0.187	0.114	0.102	0.117	0.127
TG	0.187	0.088	0.104	0.105	0.107	0.234	0.138	0.113	0.111	0.125

Table 2.6: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR; $\rho_1 = .25$ and $\rho_2 = .75$.

The results in the tables indicate that the complete pairs analysis performs very well under MCAR missingness, and even under MAR missingness in the normal case. This may be due to the fact that the probability of missingness in the second time point is symmetric around zero as a function of the values of the first time point. However, when the joint distribution is gamma, the probability of missingness is no longer symmetric and the procedure is liberal especially for the test of no main time effect.

When the groups have equal sample sizes and $\rho_1 = \rho_2$ the achieved α-level of the likelihood ratio test is satisfactory with both normal and gamma data. However, when the group sizes are not equal, the results are more liberal in all of the cases; this behavior is more severe in cases where the joint distribution is gamma, when $\rho_1 \neq \rho_2$ or when the data is MCR.

The proposed nonparametric test has satisfactory Type I error rate in all cases, for at least one of the bandwidth values. In all MCAR cases and all normal distribution cases the performance of the test is pretty robust to bandwidth selection since bandwidths of .5, 1 or 2 perform well for a variety of sample sizes, correlations, and probabilities of missingness. In the MCR cases with gamma distribution, however, the bandwidth of .5 seems to give consistently better results.

The power simulations, which are summarized in the following figures, are all for the balanced case with $N_1 = N_2 = 50$, and bandwidth of .5 for the proposed test. Figures 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6 pertain to the MCAR cases, while Figures 2.7, 2.8, 2.9, 2.10 pertain to the MCR cases. It can be seen that the power of the complete pairs analysis is always lower than the power of the proposed nonparametric test and the difference in the performance of the two is larger in the MCR gamma cases (Figures 2.8, 2.10). In all normal cases, the power of the proposed nonparametric test is almost identical to that of the likelihood ratio test (Figures 2.1, 2.3, 2.5, 2.7, and 2.9). On the other hand, the proposed test outperforms the likelihood ratio test in all gamma cases (Figures 2.2, 2.4, 2.6, 2.8, and 2.10).

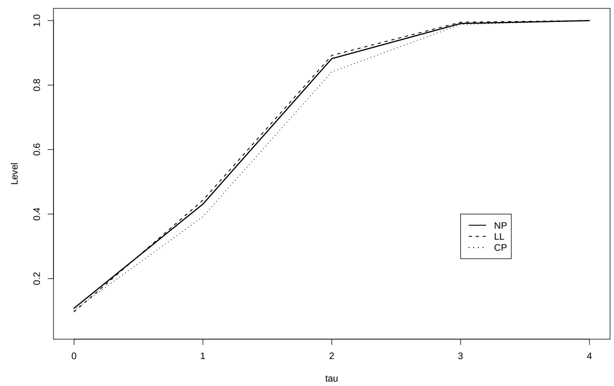

Figure 2.1: Normal MCAR case, $\rho_1 = \rho_2 = .75$, and $p_1 = p_2 = .3$.

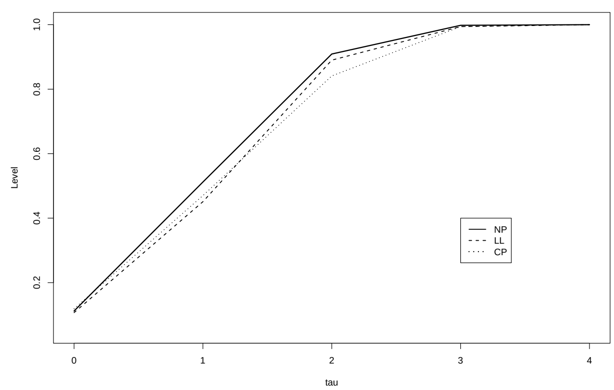

Figure 2.2: Gamma MCAR case, $\rho_1 = \rho_2 = .75$, and $p_1 = p_2 = .3$.

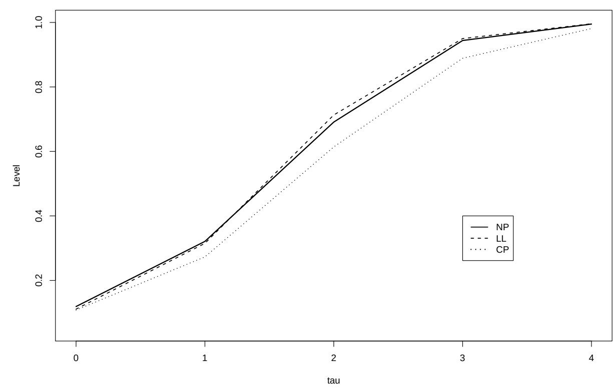

Figure 2.3: Normal MCAR case, $\rho_1 = .25,\ \rho_2 = .75$, and $p_1 = p_2 = .3$.

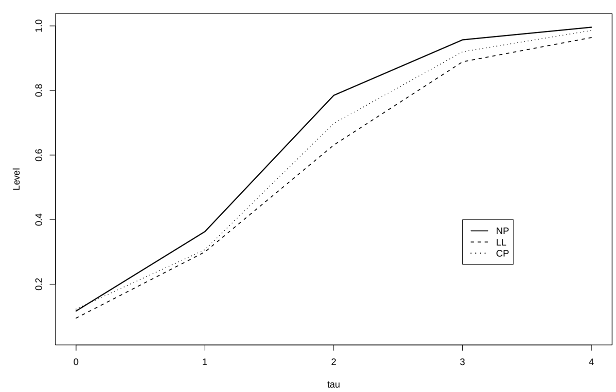

Figure 2.4: Gamma MCAR case, $\rho_1 = .25,\ \rho_2 = .75$, and $p_1 = p_2 = .3$.

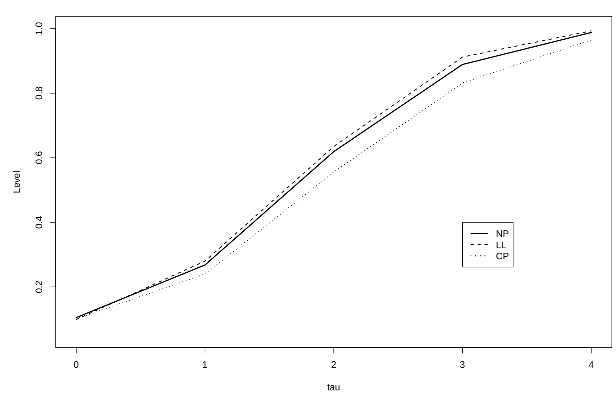

Figure 2.5: Normal MCAR case, $\rho_1 = .25,\ \rho_2 = .25$, and $p_1 = .1,\ p_2 = .3$.

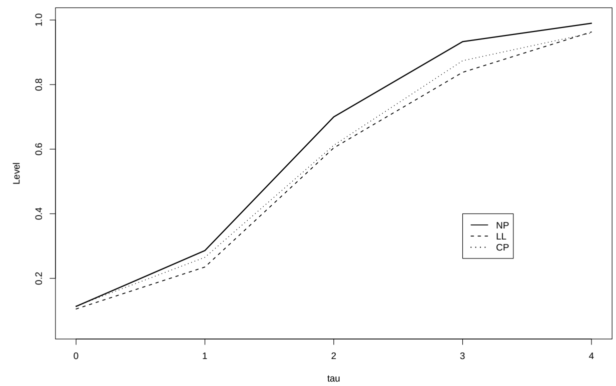

Figure 2.6: Gamma MCAR case, $\rho_1 = .25,\ \rho_2 = .25$, and $p_1 = .1,\ p_2 = .3$.

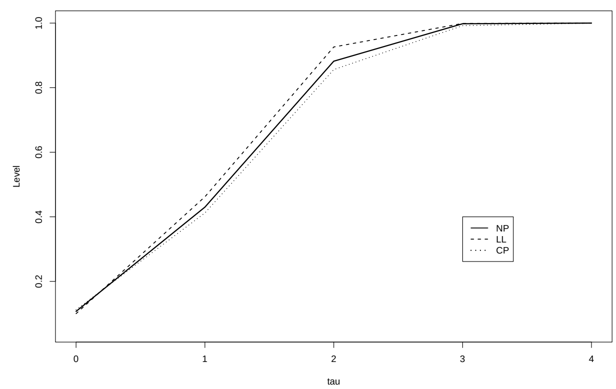

Figure 2.7: Normal MCR case, $\rho_1 = \rho_2 = .75$.

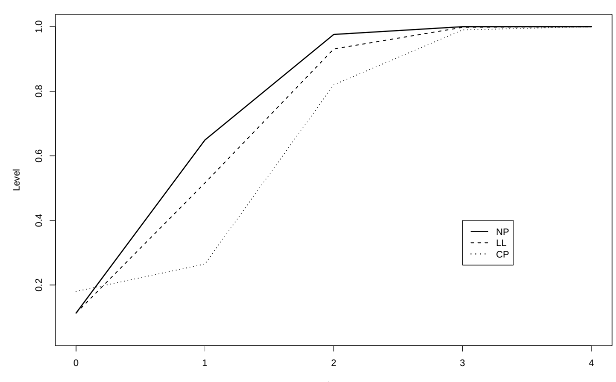

Figure 2.8: Gamma MCR case, $\rho_1 = \rho_2 = .75$.

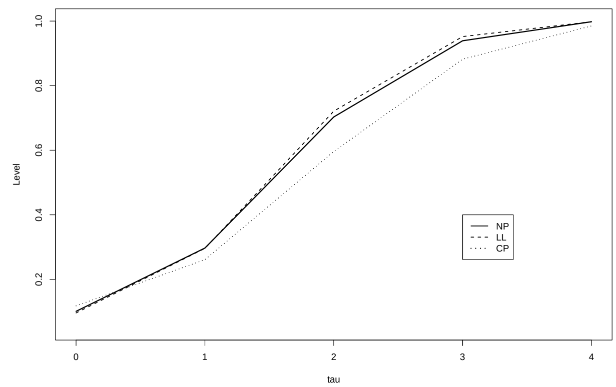

Figure 2.9: Normal MCR case, $\rho_1 = .25,\ \rho_2 = .75$.

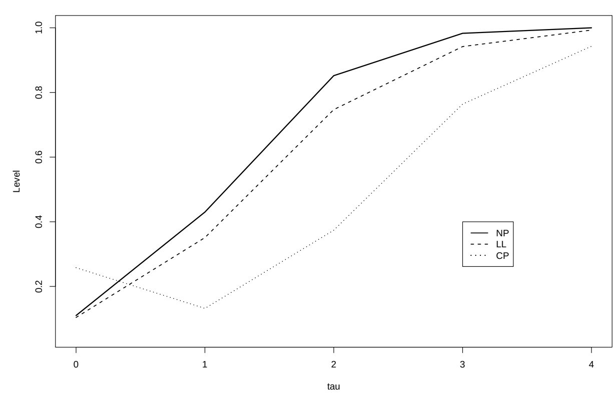

Figure 2.10: Gamma MCR case with, $\rho_1 = .25,\ \rho_2 = .75$.

2.5 Data analysis

As a practical application we will consider a study of incarcerated boys in Michigan juvenile correction institutions by Gold and Osgood (1992). The boys participated in a series of interviews trying to asses the institutional satisfaction and measure delinquent values. The interviews took place within 10 days of their arrival, after being there four months, just before leaving, six months after leaving and 36 months after leaving. During the institutional stay the response rates remained high, with 99%, 96%, 91% for the first three interviews, and 75%, 60% for the later two interviews after the boys returned to the community.

Our analysis focuses on analyzing a measure of delinquent values obtained just before leaving the institution and six months later. This pair of time points has sufficient missing data to provide good illustration of our method and it also addresses the question of whether the apparent positive change during incarceration maintains after reentering the community. The measure of delinquency is composed of eight items asking how much the respondent would admire youths who engage in delinquent behaviors versus conventionally approved behaviors. The measure of delinquency variable appears left skewed (see Figure 2.11), reflecting disapproval of delinquent acts.

In addition to the time effect, it is also of interest to study the effect of four other factors on the response: length of the most resent placement (ranges from four months to two years), type of institution (state or private), number of prior placements (ranges from zero to four) and race (African American or not). In order to have sufficient number of boys in each factor level combination we dichotomized the length of the most resent placement (up to a year or more than a year) and the number of prior placements (0, > 0). Moreover it turned out that the factor length of most recent stay has absolutely no effect. This is so in

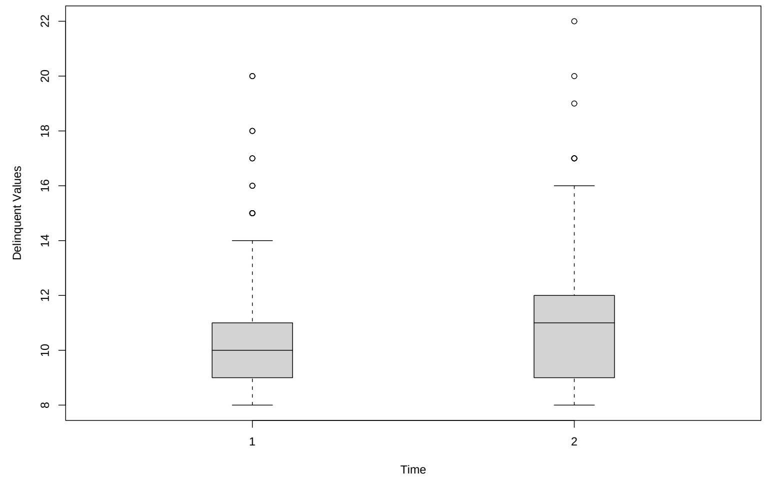

Figure 2.11: Box plots of the delinquent values of the first and the second time points we consider in our analysis.

all possible models, i.e. combinations of factors, and all methods of analysis we consider. Thus the final analysis considers only three binary factors, resulting in $a = 8$ factor level combinations, in addition to the time factor.

Out of the 338 pairs of observations 26 have both measurements missing, 6 have the first measurement missing and 60 have the second one missing. Table 2.7 summarizes the sample sizes in each of the eight factor level combinations.

Table 2.7: *Sample sizes for each factor level combination.*

i	Institution	Afr. American	Placements	N_i	n_{i1}	n_{i2}
1	State	No	0	43	36	34
2			≥ 1	35	26	22
3		Yes	0	32	27	23
4			≥ 1	17	15	14
5	Private	No	0	43	40	33
6			≥ 1	69	66	53
7		Yes	0	73	72	53
8			≥ 1	26	24	20
				312	306	252

In our analysis we considered four tests: the normal based likelihood-ratio test, the nonparametric complete-cases test and our proposed test with two different bandwidths, and normal kernel. We also considered the Epanechikov kernel but the results were very similar. The bandwidth values we used are $b_i = \nu s_p$, $i = 1, \ldots, 8$, where s_p denotes the pooled, across the factor level combinations, standard deviation of the observed data and $\nu = .5$ and 1. The analysis for the likelihood-ratio test was performed using the nmle package in R which analyzes linear mixed effects according to Laird and Ware (1982). For more details see Pinheiro and Bates (2000). Table 2.8 gives p-values for the tests of main effects and interactions for all analyses.

Effects	LR	CP	$NP_{.5}$	NP_1
Institution (I)	0.342	0.061	0.083	0.061
Afr. American (A)	0.886	0.368	0.354	0.410
Prev. Placements (P)	0.038	0.306	0.196	0.219
Time (T)	$< .001$	$< .001$	$< .001$	$< .001$
(IA)	0.180	0.055	0.024	0.022
(IP)	0.090	0.406	0.346	0.380
(IT)	0.115	0.750	0.858	0.747
(AP)	0.423	0.104	0.149	0.151
(AT)	0.906	0.986	0.793	0.877
(PT)	0.193	0.553	0.444	0.371
(IAP)	0.222	0.096	0.069	0.056
(IAT)	0.767	0.809	0.963	0.887
(IPT)	0.756	0.904	0.870	0.940
(APT)	0.943	0.479	0.701	0.664
(IAPT)	0.201	0.466	0.307	0.336

Table 2.8: p-values for main effects and interactions using different methods. LR, CP, and NP_ν denote the likelihood ratio test, the complete cases test and the proposed nonparametric test with $b_i = \nu s_p$, $i = 1, \ldots, 8$. (See text for details.)

The main time effect is highly significant with all methods. This is clear evidence that the measure of delinquent values increased after the boys left the institutions and rejoined life in their home communities. The methods are also in agreement regarding the non-significance of the main race effect, and most of the interactions. The main effect of the number of previous placements appears significant according to the LR test but not according to the nonparametric tests, while the main institution effect is significant at the 0.1 level according to the nonparametric tests. The nonparametric methods, particularly the proposed, indicate significant interaction between race and institution. This significance could have practical implications. Of practical interest might also be the moderate significance, according to the nonparametric methods, of the institution-race-placements interaction, as well as the moderate significance, according to the LR method of the institution-placements interaction. The above differences between the normal-based and nonparametric analyses may be due to the skewness of the response and the unbalancedness of the design.

2.6 Proofs

The proofs will be presented under the following technical assumptions.

Assumption A1 (i) As $N \to \infty$, N_i/N stay bounded away from zero for all $i = 1, \ldots, I$.

(ii) If Y_{ij}^0 is continuous, the corresponding bandwidth sequence b_{ij}, see (2.2.10), satisfies $Nb_{ij}^4 \to 0$ and $Nb_{ij}^{3+2\delta}(\log N)^{-1} \to \infty$, as $N \to \infty$, for some $\delta > 0$.

(iii) If Y_{ij}^0 is discrete, b_{ij} is only required to converge to zero.

Assumption A2 (i) The probability density function K is symmetric and has compact support.

(ii) The probability density function K has bounded second derivative and satisfies $\int uK(u)du = 0$.

Assumption A3 (i) If Y_{ij}^0 is discrete, then

(a) for each $j = 1, 2$, the set $\mathcal{C}_{ij}$ of mass points or atoms of the distribution of Y_{ij}^0 is the same for all i;

(b) the conditional probability mass function $f_{ij}(y|\Delta_{i1}\Delta_{i2} = 1)$ of Y_{ij}^0 given $\Delta_{i1}\Delta_{i2} = 1$ satisfies $\min\{f_{ij}(y|\Delta_{i1}\Delta_{i2} = 1); y \in \mathcal{C}_{ij}\} > 0$, for all i, j.

(ii) If Y_{ij}^0 is continuous, then

(a) the support $\mathcal{S}_{ij}$ of Y_{ij}^0 is bounded for all i, j;

(b) the conditional density $f_{ij}(y|\Delta_{i1}\Delta_{i2} = 1)$ of Y_{ij}^0 given $\Delta_{i1}\Delta_{i2} = 1$ satisfies $\inf\{f_{ij}(y|\Delta_{i1}\Delta_{i2} = 1); y \in \mathcal{S}_{ij}\} > 0$, for all i, j.

(c) Let $\widetilde{F}_{i2}(y)$ and $F_{i1}^*(y|Y_{i2k})$ denote $F_{i2}\{y|\Delta_{i2}(1 - \Delta_{i1}) = 1\}$ and $F_{i1}\{y|Y_{i2}^0 = Y_{i2k}, \Delta_{i2}(1 - \Delta_{i1}) = 1\}$, respectively. The first two derivatives $\dot{F}_{i1}^*\{y|\widetilde{F}_{i2}^{-1}(u)\}$ and $\ddot{F}_{i1}^*\{y|\widetilde{F}_{i2}^{-1}(u)\}$ of $F_{i1}^*\{y|\widetilde{F}_{i2}^{-1}(u)\}$ with respect to u exist and are bounded for all u, y and i. Similarly for $\widetilde{F}_{i1}(y)$ and $F_{i2}^*(y|Y_{i1k})$.

(d) Let $g_{i2}(y) = f_{i2}\{y|\Delta_{i2}(1 - \Delta_{i1})\}$. The first two derivatives of $g_{i2}(y)$ exist and are bounded. Similarly for $g_{i1}(y)$.

Proof of Proposition 2.3.1. It suffices to show this for $j = 1$. For simplicity in notation we will let $\widehat{F}_{i1}^*(y|Y_{i2k})$ and $F_{i1}^*(y|Y_{i2k})$ denote $\widehat{F}_{i1}(y|Y_{i2}^0 = Y_{i2k}, \Delta_{i2} = 1, \Delta_{i1} = 0)$ and $F_{i1}(y|Y_{i2}^0 = Y_{i2k}, \Delta_{i2} = 1, \Delta_{i1} = 0)$, respectively. Note that by assumption (2.2.2) we have $F_{i1}^*(y|Y_{i2k}) =$

$F_{i1}\left(y|Y_{i2}^0 = Y_{i2k}, \Delta_{i2} = 1\right)$. Using (2.2.12) and (2.2.9) we write

$$\begin{aligned}
& \hat{F}_{i1}(y) - F_{i1}(y) \\
= \; & \frac{1}{n_i}\sum_{k=1}^{N_i}\{\Delta_{i1k}c(Y_{i1k}^0, y) + \Delta_{i2k}(1-\Delta_{i1k})\widehat{F}_{i1}^*\left(y|Y_{i2k}\right)\} - F_{i1}(y) \\
= \; & \frac{1}{n_i}\sum_{k=1}^{N_i}\{\Delta_{i1k} + \Delta_{i2k}(1-\Delta_{i1k})\}\{c(Y_{i1k}^0, y) - F_{i1}(y)\} \\
& + \frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\{F_{i1}^*(y|Y_{i2k}) - c(Y_{i1k}^0, y)\} \\
& + \frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left[\sum_{l=1}^{N_i} w_{i2l}(Y_{i2k})\{c(Y_{i1l}^0, y) - F_{i1}^*(y|Y_{i2l})\}\right] \\
& + \frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left[\sum_{l=1}^{N_i} w_{i2l}(Y_{i2k})\{F_{i1}^*(y|Y_{i2l}) - F_{i1}^*(y|Y_{i2k})\}\right]. \quad (2.6.1)
\end{aligned}$$

Note that the first term on the right hand side of (2.6.1) is centered by the fact that $\{\Delta_{i1k} + \Delta_{i2k}(1-\Delta_{i1k}) = 1\} = \{\Delta_{i1k} = 1 \text{ or } \Delta_{i2k} = 1\}$ and assumption (2.2.1) according to which $[\Delta_{i1k} = 1 \text{ or } \Delta_{i2k} = 1]$ is independent from (Y_{i1k}^0, Y_{i2k}^0). Also, the second term is easily seen to be centered by conditioning on $Y_{i2}^0 = Y_{i2k}, \Delta_{i2k} = 1, \Delta_{i1k} = 0$. The third term on the right hand side of (2.6.1) is seen to be centered by conditioning on those (Y_{i2l}, Δ_{i2l}) pairs for which $\Delta_{i2l} = 1$, since, by assumption (2.2.2) Y_{i1l}^0 is independent from Δ_{i1l}, conditionally on Y_{i2l} and $\Delta_{i2l} = 1$. Thus, it is centered also unconditionally. The final term on the right hand side of (2.6.1) is the bias term.

Using the integration by parts formula found, e.g. in Hewitt and Stromberg (1969, p. 419), we have

$$\int \left(\widehat{H} - H\right) d\left(\widehat{F}_{i1} - F_{i1}\right) = -\int \left(\widehat{F}_{i1} - F_{i1}\right) d\left(\widehat{H} - H\right),$$

which we can write as

$$-\sum_{i'=1}^{I}\sum_{j=1}^{2}\frac{n_{ij}}{n_{..}}\int \left(\widehat{F}_{i1} - F_{i1}\right) d\left(\widehat{H}_{i'j} - H_{i'j}\right)$$

where $\widehat{H}_{i'j}(y) = (1/n_{i'j})\sum_{k=1}^{N_i'} \Delta_{i'jk} c(Y_{i'jk}, y)$ and $E(\widehat{H}_{i'j}) = H_{i'j}$. Thus in order to show that $N^{1/2}\int \left(\widehat{H} - H\right) d\left(\widehat{F}_{i1} - F_{i1}\right) \to 0$, it suffices to show that $N^{1/2}\int \left(\widehat{F}_{i1} - F_{i1}\right) d\left(\widehat{H}_{i'j} - H_{i'j}\right) \to 0$ for each i', j. By (2.6.1), this entails showing

$$\frac{1}{n_i}\sum_{k=1}^{N_i}\{\Delta_{i1k} + \Delta_{i2k}(1-\Delta_{i1k})\}\int \left\{c(Y_{i1k}^0, y) - F_{i1}(y)\right\} d\left\{\widehat{H}_{i'j}(y) - H_{i'j}(y)\right\}$$
$$= o_p\left(N^{-1/2}\right) \qquad (2.6.2)$$

$$\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\int \left\{F_{i1}^*(y|Y_{i2k}) - c(Y_{i1k}^0, y)\right\} d\left\{\widehat{H}_{i'j}(y) - H_{i'j}(y)\right\}$$
$$= o_p\left(N^{-1/2}\right) \qquad (2.6.3)$$

$$\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\sum_{l=1}^{N_i} w_{i2l}(Y_{i2k})\int \left\{c(Y_{i1l}^0, y) - F_{i1}^*(y|Y_{i2l})\right\}$$
$$d\left\{\widehat{H}_{i'j}(y) - H_{i'j}(y)\right\} = o_p\left(N^{-1/2}\right) \qquad (2.6.4)$$

$$\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\sum_{l=1}^{N_i} w_{i2l}(Y_{i2k})\int \left\{F_{i1}^*(y|Y_{i2l}) - F_{i1}^*(y|Y_{i2k})\right\}$$
$$d\left\{\widehat{H}_{i'j}(y) - H_{i'j}(y)\right\} = o_p\left(N^{-1/2}\right). \qquad (2.6.5)$$

That (2.6.2) and (2.6.3) hold follows by straightforward variance calculations. To show (2.6.4) and (2.6.5) we first note that the denominator, $n_{ci}\widehat{f}_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2} = 1)$ of the weight function can be replaced with $n_{ci}f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2} = 1)$. In what follows we assume that this replacement has been made, and we write $\tilde{w}_{ijk}(x)$ for the weights resulting from this substitution. We will first show (2.6.5). Set

$$\begin{aligned} q_{kl}^{i1}(Y_{i'jm}) &= \Delta_{i'jm}F_{i1}^*(Y_{i'jm}|Y_{i2l}) - \Delta_{i'jm}F_{i1}^*(Y_{i'jm}|Y_{i2k}) \\ &\quad - \int \left\{F_{i1}^*(y|Y_{i2l}) - F_{i1}^*(y|Y_{i2k})\right\} dH_{i'j}(y), \end{aligned}$$

and let T_1 denote the left-hand side of (2.6.5). Noting that $\mathrm{E}\{q_{kl}^{i1}(Y_{i'jm})|\Delta_{i'jm} = 1, \mathbf{Z}_k, \mathbf{Z}_l\} = 0$ if $k, l \neq m$, we have that $E(NT_1^2)$ is equal to

$$
\begin{aligned}
& NE\left\{\frac{1}{n_i^2}\sum_{k_1,k_2}^{N_i}\Delta_{i2k_1}(1-\Delta_{i1k_1})\Delta_{i2k_2}(1-\Delta_{i1k_2})\sum_{l_1,l_2}^{N_i}\tilde{w}_{i2l_1}(Y_{i2k_1})\tilde{w}_{i2l_2}(Y_{i2k_2})\right.\\
& \left.\times\frac{1}{n_{i'j}^2}\sum_{m_1,m_2}^{n_{i'j}} q_{k_1l_1}^{i1}(Y_{i'jm_1})q_{k_2l_2}^{i1}(Y_{i'jm_2})\right\}\\
= \;& NE\left[\frac{1}{n_i^2}\sum_{k_1,k_2}^{N_i}\Delta_{i2k_1}(1-\Delta_{i1k_1})\Delta_{i2k_2}(1-\Delta_{i1k_2})\sum_{l_1,l_2}^{N_i}\tilde{w}_{i2l_1}(Y_{i2k_1})\tilde{w}_{i2l_2}(Y_{i2k_2})\right.\\
& \left.\times\frac{1}{n_{i'j}^2}\sum_{m_1,m_2}^{N_i} E\left\{q_{k_1l_1}^{i1}(Y_{i'jm_1})q_{k_2l_2}^{i1}(Y_{i'jm_2})|\mathbf{Z}_{k_1},\mathbf{Z}_{k_2},\mathbf{Z}_{l_1},\mathbf{Z}_{l_2}\right\}\right]\\
= \;& NE\left(\frac{1}{n_i^2}\sum_{k_1,k_2}^{N_i}\Delta_{i2k_1}(1-\Delta_{i1k_1})\Delta_{i2k_2}(1-\Delta_{i1k_2})\sum_{l_1,l_2}^{N_i}\tilde{w}_{i2l_1}(Y_{i2k_1})\tilde{w}_{i2l_2}(Y_{i2k_2})\right.\\
& \left.\times\frac{1}{n_{i'j}^2}\left[\sum_{m=1}^{N_i} E\left\{q_{k_1l_1}^{i1}(Y_{i'jm})q_{k_2l_2}^{i1}(Y_{i'jm})|\mathbf{Z}_{k_1},\mathbf{Z}_{k_2},\mathbf{Z}_{l_1},\mathbf{Z}_{l_2}\right\} + R_{k_1,k_2,l_1,l_2}\right]\right)
\end{aligned}
$$

where

$$R_{k_1,k_2,l_1,l_2} = \sum_{m_1,m_2} q_{k_1l_1}^{i1}(Y_{i'jm_1})q_{k_2l_2}^{i1}(Y_{i'jm_2})I\left\{m_1, m_2 \in \{k_1, k_2, l_1, l_2\},\ m_1 \neq m_2\right\},$$

are uniformly bounded random variables. With a one-term Taylor expansion and using assumption A3(ii)(c), it follows that $|q_{k_1l_1}^{i1}(Y_{i'jm})q_{k_2l_2}^{i1}(Y_{i'jm})| \leq C|Y_{i2k_1} - Y_{i2l_1}||Y_{i2k_2} - Y_{i2l_2}|$, for some constant C. Thus, taking into account the proximity of Y_{i2k_1} and Y_{i2l_1}, which is implied by the weight functions $\tilde{w}_{i2l_1}(Y_{i2k_1})$, $\tilde{w}_{i2l_2}(Y_{i2k_2})$, it is easily seen that $E(NT_1^2) = O(b_{i2}^2)$. This completes the proof of (2.6.5).

Let T_2 denote the left hand side of (2.6.4). Write

$$
\begin{aligned}
T_2 &= \frac{1}{n_i}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\Bigg[\Delta_{i2k}(1-\Delta_{i1k})\tilde{w}_{i2l}(Y_{i2k})\frac{1}{n_{i'j}}\sum_{m=1}^{N_{i'}}\Delta_{i'jm}\Big\{\, c(Y_{i1l},Y_{i'jm}) \\
&\quad -F_{i1}^*(Y_{i'jm}|Y_{i2l}) - 1 + H_{i'j}(Y_{i1l}) + \int F_{i1}^*(y|Y_{i2l})dH_{i'j}(y)\Big\}\Bigg] \\
&= \sum_{l=1}^{N_i}\left[\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left\{\tilde{w}_{i2l}(Y_{i2k}) - \int \tilde{w}_{i2l}(x)g_{i2}(x)dx\right\}\right] \\
&\qquad \times\Bigg[\frac{1}{n_{i'j}}\sum_{m=1}^{N_{i'}}\Delta_{i'jm}\Big\{c(Y_{i1l},Y_{i'jm}) - F_{i1}^*(Y_{i'jm}|Y_{i2l}) - 1 + H_{i'j}(Y_{i1l}) \\
&\qquad + \int F_{i1}^*(y|Y_{i2l})dH_{i'j}(y)\Big\}\Bigg] \\
&\quad + \frac{1}{n_{i'j}}\sum_{m=1}^{N_{i'}}\sum_{l=1}^{N_i}\frac{1}{n_{ci}}\frac{g_{i2}(Y_{i2l})\Delta_{i1l}\Delta_{i2l}}{f_{i2}(Y_{i2l}|\Delta_{i1}\Delta_{i2}=1)}\Delta_{i'jm}\Big\{\, c(Y_{i1l},Y_{i'jm}) \\
&\qquad -F_{i1}^*(Y_{i'jm}|Y_{i2l}) - 1 + H_{i'j}(Y_{i1l}) + \int F_{i1}^*(y|Y_{i2l})dH_{i'j}(y)\Big\} \\
&\quad + O_p\left(\frac{b_{i2}^2}{n_{i1}}\right),
\end{aligned}
\tag{2.6.6}
$$

where $g_{i2}(x)$ is defined in assumption A3(d) and the second equality used the relation

$$
\int \tilde{w}_{i2l}(x)g_{i2}(x)dx = \frac{1}{n_{ci}}\left\{\frac{g_{i2}(Y_{i2l})\Delta_{i1l}\Delta_{i2l}}{f_{i2}(Y_{i2l}|\Delta_{i1}\Delta_{i2}=1)} + O(b_{i2}^2)\right\},
$$

which follows by a change of variable and a two term Taylor expansion. We will first show that

$$
\begin{aligned}
&\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left\{\tilde{w}_{i2l}(Y_{i2k}) - \int \tilde{w}_{i2l}(x)g_{i2}(x)dx\right\} \\
&\qquad = o_p\{(n_{ci}b_{i2})^{-1}N^{-1/2}\},
\end{aligned}
\tag{2.6.7}
$$

$$
\begin{aligned}
&\frac{1}{n_{i'j}}\sum_{m=1}^{N_{i'}}\Delta_{i'jm}\Big\{\, c(Y_{i1l},Y_{i'jm}) - F_{i1}^*(Y_{i'jm}|Y_{i2l}) - 1 + H_{i'j}(Y_{i1l}) \\
&\qquad + \int F_{i1}^*(y|Y_{i2l})dH_{i'j}(y)\Big\} = o_p\{(\log N)^{1/2}N^{-1/2}\}
\end{aligned}
\tag{2.6.8}
$$

uniformly in l which will imply that the first term on the right hand side of (2.6.6) is $O_p\{(\log N)^{1/2}/(Nb_{i2})\} = o_p(N^{-1/2})$. To show relation (2.6.7) we use Bernstein's inequality Uspensky (1937). to obtain

$$\begin{aligned}
&P\left[\max_{1\le l\le N_i}\left|\frac{n_{ci}b_{i2}N^{1/2}}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left\{\tilde{w}_{i2l}(Y_{i2k})-\int\tilde{w}_{i2l}(x)g_{i2}(x)dx\right\}\right|>\epsilon\right]\\
&\quad= P\left[\max_{1\le l\le N_i}\left|\frac{N^{1/2}}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left\{w^*_{i2l}(Y_{i2k})-\int w^*_{i2l}(x)g_{i2}(x)dx\right\}\right|>\epsilon\right]\\
&\quad\le \sum_{l=1}^{N_i}P\left[\left|\frac{N^{1/2}}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left\{w^*_{i2l}(Y_{i2k})-\int w^*_{i2l}(x)g_{i2}(x)dx\right\}\right|>\epsilon\right]\\
&\qquad\le N_i 2\exp\left(-\frac{N^{-1}n_i^2\epsilon^2}{N_ib_{i2}C_1+C_2N^{-1/2}n_i\epsilon}\right)\to 0,\quad \text{as}\quad N\to\infty,
\end{aligned}$$

where $w^*_{i2l}(x) = n_{ci}\tilde{w}_{i2l}(x) = \Delta_{i1l}\Delta_{i2l}K_{b_{i2}}(Y_{i2l}-x)/f_{i2}(x|\Delta_{i1}\Delta_{i2}=1)$, and C_1, C_2 are some positive constants. Relation (2.6.8) is shown by a similar application of Bernstein's inequality.

The second term of the right hand side of (2.6.6) can be written in the form of a U-statistic, namely as $(n_{i'j}n_{ci})^{-1}\sum_m\sum_l\phi(\mathbf{Z}_{il},\mathbf{Z}_{i'm})$, where the definition of the kernel ϕ is implicit. It is easily verified that $E\{\phi(\mathbf{Z}_{il},\mathbf{Z}_{i'm})|\mathbf{Z}_{il}\} = E\{\phi(\mathbf{Z}_{il},\mathbf{Z}_{i'm})|\mathbf{Z}_{i'm}\} = 0$. Thus, from the theory of U-statistics it follows that this term is $O_p(N^{-1})$ and so $o_P(N^{-1/2})$. This completes the proof of (2.6.6) and of Proposition 2.3.1.

Proof of Theorem 2.3.2. According to (2.6.1), the theorem will be established by showing

$$\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\sum_{l=1}^{N_i}w_{i2l}(Y_{i2k})\left[H(Y_{i1l}^0)-E_{i1}^*\{H(Y_{i1}^0)|Y_{i2l}\}\right]$$
$$=\frac{2}{n_{ci}}\sum_{k=1}^{N_i}H_{1,1}(\mathbf{Z}_{ik})+o_p\left(N^{-1/2}\right) \quad (2.6.9)$$

$$\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\sum_{l=1}^{N_i}w_{i2l}(Y_{i2k})\int H(y)\ d\{F_{i1}^*(y|Y_{i2l})-F_{i1}^*(y|Y_{i2k})\}$$
$$=o_p(N^{-1/2}). \quad (2.6.10)$$

First (2.6.10) follows by arguments similar to those used in (2.6.5). Next we show (2.6.9):

$$\begin{aligned}
&\frac{1}{n_i}\sum_{k=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\sum_{l=1}^{N_i}w_{i2l}(Y_{i2k})\left[H(Y_{i1l}^0)-E_{i1}^*\{H(Y_{i1}^0)|Y_{i2l}\}\right]\\
=\ &\frac{1}{n_i}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\frac{\Delta_{i2k}(1-\Delta_{i1k})}{n_{ci}}\left\{\frac{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2}=1)}{\frac{1}{n_{ci}}\sum_m K_{b_{i2}}(Y_{i2k}-Y_{2jm})\Delta_{2jm}\Delta_{1jm}}-1+1\right\}\\
&\times\frac{K_{b_{i2}}(Y_{i2k}-Y_{i2l})\Delta_{i2l}\Delta_{i1l}}{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2}=1)}\left[H(Y_{i1l}^0)-E_{i1}^*\{H(Y_{i1}^0)|Y_{i2l}\}\right]\\
=\ &\frac{1}{n_i}\frac{1}{n_{ci}}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\left\{\frac{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2}=1)}{\frac{1}{n_{ci}}\sum_m K_{b_{i2}}(Y_{i2k}-Y_{2jm})\Delta_{2jm}\Delta_{1jm}}-1\right\}\\
&\times\frac{K_{b_{i2}}(Y_{i2k}-Y_{i2l})\Delta_{i2l}\Delta_{i1l}}{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2}=1)}\left[H(Y_{i1l}^0)-E_{i1}^*\{H(Y_{i1}^0)|Y_{i2l}\}\right]\\
+\ &\frac{1}{n_i}\frac{1}{n_{ci}}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\Delta_{i2k}(1-\Delta_{i1k})\frac{K_{b_{i2}}(Y_{i2k}-Y_{i2l})\Delta_{i2l}\Delta_{i1l}}{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2}=1)}\\
&\times\left[H(Y_{i1l}^0)-E_{i1}^*\{H(Y_{i1}^0)|Y_{i2l}\}\right].
\end{aligned}$$

Using the notation introduced in the statement of the theorem, the first and second terms on the right hand side of the above relation are, respectively,

$$\frac{1}{n_i}\frac{1}{n_{ci}}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\widetilde{H}_1(\mathbf{Z}_{ik},\mathbf{Z}_{il})\left\{\frac{f_{i2}(Y_{i2k}|\Delta_{i1}\Delta_{i2}=1)}{\frac{1}{n_{ci}}\sum_m K_{b_{i2}}(Y_{i2k}-Y_{2jm})\Delta_{2jm}\Delta_{1jm}}-1\right\}, \quad (2.6.11)$$

$$\frac{1}{n_i}\frac{1}{n_{ci}}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\widetilde{H}_1(\mathbf{Z}_{ik},\mathbf{Z}_{il}). \quad (2.6.12)$$

Direct variance calculations using the fact that $\widetilde{H}_1(\mathbf{Z}_{ik},\mathbf{Z}_{il})$ has zero conditional expectation given the pairs (Y_{i2k},Δ_{i2k}) for which $\Delta_{i2k}=1$, and assumption A3(i)(b) for the discrete case, or A3(ii)(b) together with the uniform consistency of the kernel density estimator Silverman (1978) in the continuous case, show that the term in (2.6.11) is $o_p\left(N_i^{-1/2}\right)$, conditionally on the pairs (Y_{i2k},Δ_{i2k}) for which $\Delta_{i2k}=1$, and thus also unconditionally. Consider now (2.6.12), and recall the notation $H_1(\mathbf{Z}_{ik},\mathbf{Z}_{il})$ introduced in the statement of the theorem. Keeping in mind that $\widetilde{H}_1(\mathbf{Z}_{ik},\mathbf{Z}_{il})=0$ if $k=l$, and also that we are essentially working with the n_i pairs having at least one non-missing observation (thus, the summation indices in (2.6.12) extend to N_i only in order to avoid awkward renumbering of the data), (2.6.12) equals

$$\begin{aligned}\frac{1}{n_i}\frac{1}{n_{ci}}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}H_1(\mathbf{Z}_{ik},\mathbf{Z}_{il}) &= \frac{1}{n_i}\frac{1}{n_{ci}}n_i(n_i-1)\left\{\frac{2}{n_i(n_i-1)}\sum_{k<l}H_1(\mathbf{Z}_{ik},\mathbf{Z}_{il})\right\}\\ = \frac{n_i-1}{n_{ci}}\left\{\frac{2}{n_i}\sum_k H_{1,1}(\mathbf{Z}_{ik})+o_p\left(n_i^{-1/2}\right)\right\} &= \frac{2}{n_{ci}}\sum_{k=1}^{N_i}H_{1,1}(\mathbf{Z}_{ik})+o_p\left(N^{-1/2}\right),\end{aligned}$$

where the second equality above follows from standard theory for U-statistics. This completes the proof of the theorem.

Chapter 3

Factorial Designs with MCR Repeated Measurements

3.1 Introduction and Notation

In this chapter we extend the ideas introduced in Chapter 2 for cases were there are more than two repeated measurements per subject. We propose a general notation to describe different missingness patterns and extend the MCR assumption for arbitrary number of time points. With the testing procedures prosed here we reanalyze the delinquency data using the information from four interviews for each boy.

Let $(Y^0_{i1k}, \ldots, Y^0_{iJk})$, $k = 1, \ldots, N_i$, be independent replications of the variables of interest $(Y^0_{i1}, \ldots, Y^0_{iJ})$, where the index i denotes the factor-level combinations, excluding the factor time which is denoted by the second subscript. For example, in the incarcerated boys application, if we consider only the factors prior placement (yes or no) and length of most recent stay, their are four factor levels enumerated as 1: yes & up to a year, 2: yes & more than a year, 3: no & up to a year, 4: no & more than a year, and $(Y^0_{i1k}, \ldots, Y^0_{i4k})$ are the measures of delinquency in the four interviews for the k-th boy with factor-level combination i. Since some of these variables are missing, for each subject we observe

$$\mathbf{Z}_{ik} = (Y_{i1k}, \Delta_{i1k}, \ldots, Y_{iJk}, \Delta_{iJk}), \quad k = 1, \ldots, N_i, \tag{3.1.1}$$

where $\Delta_{ijk} = 1$ if Y^0_{ijk} is observed, in which case $Y_{ijk} = Y^0_{ijk}$ and $\Delta_{ijk} = 0$ if Y^0_{ijk} is missing, in which case Y_{ijk} is arbitrary.

Similarly as in the paired data case, we only assume

$$Y^0_{ijk} \sim F_{ij}, \text{ for all } i, j, \text{ and } k. \tag{3.1.2}$$

and hypotheses in the context of this nonparametric model are defined as Akritas and Arnold (1994). (For more details see Chapter 2.)

We are also introducing the notation for the simpler case of no factors (besides the time factor), which will be used for the discussion in Sections 3.2 and 3.3. Since there are no factors, the index i is dropped, and we have $(Y^0_{1k}, \ldots, Y^0_{Jk})$ for $k = 1 \ldots, N$, independent replications of the variables $(Y^0_1, \ldots, Y^0_J)$, where $Y^0_j \sim F_j$ for $j = 1, \ldots, J$. The hypothesis of interest in this case is

$$H_0 \;:\; F_1 = \ldots = F_J. \tag{3.1.3}$$

$$\Leftrightarrow \quad H_0 \;:\; \mathbf{CF} = 0, \tag{3.1.4}$$

where $\mathbf{F} = (F_1, \ldots, F_J)'$ and $\mathbf{C}$ is the contrast matrix of full row rank $\mathbf{C} = (\mathbf{1}_{a-1}| - \mathbf{I}_{a-1})$, for $\mathbf{1}_d = (1, \ldots, 1)'$ is a d-dimensional vector of ones, and $\mathbf{I}_d$ denotes the d-dimensional unit matrix.

Finally, similar to (3.1.1), assume we observe

$$\mathbf{Z}_k = (Y_{1k}, \Delta_{1k}, \ldots, Y_{Ik}, \Delta_{Ik}), \quad k = 1, \ldots, N, \tag{3.1.5}$$

where $\Delta_{jk} = 1$ if Y^0_{jk} is observed, in which case $Y_{jk} = Y^0_{jk}$ and $\Delta_{jk} = 0$ if Y^0_{jk} is missing, in which case Y_{jk} is arbitrary.

To test the null hypothesis (3.1.3) nonparametrically we must be able to estimate F_j nonparametrically. We will consider estimation of F_j under the general MCR assumption, that accounts for more than two time points, which will be specified later. For now we will only use the part of it that guarantees that cases whose variables are all missing contain no useful information. This implies

$$\begin{aligned} F_j(y) &= F_j(y|\text{at least one variable is observed}) \\ &= F_j(y| \cup_{s=1}^{J} [\Delta_s = 1]). \end{aligned} \tag{3.1.6}$$

The plan to do so is to first decompose the event $\cup_{s=1}^{J} [\Delta_s = 1]$ into J disjoint events each of which specifies that a particular coordinate is observed, and then use, on each of the disjoint events, the observed coordinate for the imputation. Let $S_1^j, \ldots, S_J^j$ denote the partition of $\cup_{s=1}^{J} [\Delta_s = 1]$ which will be used for estimating F_j. We will use the relation

$$\begin{aligned} F_j(y| \cup_{s=1}^{J} [\Delta_s = 1]) = & \quad F_j\left(y|S_1^j\right) P(S_1^j| \cup_{s=1}^{J} [\Delta_s = 1]) \\ & + F_j\left(y|S_2^j\right) P(S_2^j| \cup_{s=1}^{J} [\Delta_s = 1]) \\ & \vdots \\ & + F_j\left(y|S_J^j\right) P(S_J^j| \cup_{s=1}^{J} [\Delta_s = 1]). \end{aligned}$$

Clearly, the probabilities $P(S_r^j| \cup [\Delta_j = 1])$ can be estimated with n_r^j/n, for n_r^j the number of cases with missingness pattern S_r^j and n the number of cases with at least one observed variable. To describe the estimation of $F_j(y|S_r^j)$, we need to be more specific about the events S_r^j. Let $j, j_2, \ldots, j_J$ be a permutation of $\{1, \ldots, J\}$ whose first element is j and set

$$\begin{aligned} S_1^j &= [\Delta_j = 1] \\ S_2^j &= [\Delta_j = 0] \cap [\Delta_{j_2} = 1] \\ S_3^j &= [\Delta_j = 0] \cap [\Delta_{j_2} = 0] \cap [\Delta_{j_3} = 1] \\ &\vdots \\ S_J^j &= [\Delta_j = 0] \cap \ldots [\Delta_{j_{(a-1)}} = 0] \cap [\Delta_{j_J} = 1]. \end{aligned} \tag{3.1.7}$$

Note that j_r is the coordinate index which is specified to be observed on S_r^j. Since $S_1^j = [\Delta_j = 1]$, $F_j(y|S_1^j)$, is the marginal distribution of Y_j^0 given that it is observed, and so it is consistently estimated by the empirical distribution function of the observed Y_j^0's. Less obvious is the estimation of $F_j(y|S_r^j)$, $r = 2, \ldots, J$, since Y_j^0 is missing on events S_r^j, $r = 2, \ldots, J$. Basically, $F_j(y|S_r^j)$ is estimated by nonparametric imputation based on the coordinate $Y_{j_r}^0$ which is guaranteed to be observed on S_r^j.

Remark 3.1.1. *A sensible way to choose the permutation $j, j_2, \ldots, j_J$ for each j is to let j_2 be the time point closest to j (with preference to the time point on the left), j_3 to be the second closest time point to j end so on. For example, if $a = 6$, a reasonable set of permutations is the following*

j	j	j_2	j_3	j_4	j_5	j_6
1	1	2	3	4	5	6
2	2	1	3	4	5	6
3	3	2	4	1	5	6
4	4	3	5	2	6	1
5	5	4	6	3	2	1
6	6	5	4	3	2	1

Thus, for estimating F_4 we choose the permutation 4,3,5,2,6,1. This means that if Y_4^0 is missing, we first consider the set of cases where Y_3^0 is observed and perform nonparametric imputation based on it. Next we consider the set of cases where Y_5^0 is observed and perform nonparametric imputation based on it, and so on. This way, imputation is always based on the nearest observed variable.

Remark 3.1.2. *The sets* S_r^j, $r = 1, \ldots, J$ *will also be thought as a partition of the set of all subsets. For example,* S_1^j *is the subset of subjects with* $\Delta_{jk} = 1$, S_2^j *is the subset of subjects with* $\Delta_{jk} = 0$, *and* $\Delta_{j_2} = 1$, *and so on; see* (3.1.7).

The theory of this analysis under the framework of simple repeated measurement design, (i.e. only time factor) is presented in Sections 3.2 and 3.3. In Section 3.4 we extend the results to higher way factorial designs. Results from simulation studies are presented in Section 3.5 and analysis of the sociological data set introduced in the previous chapter is given in Section 3.6. Finally, in Section 3.7 we present the proofs of the main results given in this chapter.

3.2 Estimation of the marginal distribution functions

As we have seen in the introduction, based on (3.1.6) and the fact that $\{S_r^j\}_{r=1}^J$ is a partition of $\cup_{s=1}^J[\Delta_s = 1]$, we have the following decomposition of the cumulative distribution function of Y_j^0

$$F_j(y) = \sum_{r=1}^{J} F_j\left(y|S_r^j\right) P(S_r^j| \cup_{s=1}^J [\Delta_s = 1]). \tag{3.2.1}$$

As already mentioned, the probabilities $P(S_r^j|\cup[\Delta_j = 1])$ are estimated by n_r^j/n, and $F_j\left(y|S_1^j\right)$ is estimated by

$$\widehat{F}_j\left(y|S_1^j\right) = \widehat{F}_j\left(y|\Delta_j = 1\right) = \frac{1}{n_j}\sum_{k=1}^{N} \Delta_{jk} c(Y_{jk}, y) \tag{3.2.2}$$

where the function c as defined in (2.1.5) and n_j is the number of cases with Y_j^0 observed.

Since on S_r^j, $r = 2, \ldots, J$, Y_j^0 is missing, $F_j(y|S_r^j)$ will be estimated by nonparametric imputation using an appropriate set of "donor". Deciding on the set of donors essentially determines the MCR assumptions. We now define a particular set of MCR assumptions and identify the corresponding donor sets. In words, these MCR assumptions state that given the value of the nearest nonmissing observation, Y_j^0 and Δ_j are independent. Formally, we have

$$F_j(y|Y_{j_r} = v, S_r^j) = F_j(y|Y_{j_r} = v, \widetilde{S}_r^j), \quad \text{for } r = 2, \ldots, J \tag{3.2.3}$$

where S_r^j is as defined in (3.1.7) and

$$\tilde{S}_r^j = [\Delta_j = 1] \cap [\Delta_{j_2} = 0] \cap \ldots [\Delta_{j_{(r-1)}} = 0] \cap [\Delta_{j_r} = 1], \quad \text{for } r = 2, \ldots, J.$$

In other words, thinking of the S_r^j as subsets of the set of subjects (see Remark 3.1.2), $\tilde{S}_r^j$ is the set of donors for missing pattern S_r^j.

Remark 3.2.1. *An alternative sequence of donor sets might be of the form*

$$\breve{S}_r^j = [\Delta_j = 1] \cap [\Delta_{j_r} = 1], \quad \textit{for } r = 2, \ldots, J.$$

These donor sets might be preferred over the $\tilde{S}_r^j$ in cases with not very large data sets, since the donor sets $\breve{S}_r^j$ are larger than $\tilde{S}_r^j$. Even though the choice of $\breve{S}_r^j$ sets will imply stronger MCR assumptions than the one defined in (3.2.3)*, these assumptions are still weaker than the strongly ignorable MAR assumption.*

Using (3.2.3), we have that

$$\begin{aligned} F_j\left(y|S_r^j\right) &= \int F_j\left(y|Y_{j_r}^0 = v, S_r^j\right) dF_{j_r}\left(v|S_r^j\right) \\ &= \int F_j\left(y|Y_{j_r}^0 = v, \widetilde{S}_r^j\right) dF_{j_r}\left(v|S_r^j\right). \end{aligned} \tag{3.2.4}$$

Recall that j_r is the coordinate index which is specified to be observed on S_r^j, and thus $F_{j_r}(v|S_r^j)$ is readily estimated by a corresponding empirical distribution function on the observed $Y_{j_r}^0$ values from the S_r^j-cases.

Estimation of $F_j\left(y|Y_{j_r}^0 = v, \widetilde{S}_r^j\right)$ is accomplished using smoothing techniques involving a symmetric kernel function K and a bandwidth b. Basically, this conditional distribution function is estimated using the values Y_{jk} in the $\widetilde{S}_r^j$-cases, based on how "close" is Y_{j_rk} to v. More specifically, denoting $K_b(x) = K(x/b)/b$, and under some smoothing assumptions we have

$$\widehat{F}_j\left(y|Y_{j_r}^0 = v, \widetilde{S}_r^j\right) = \sum_{k=1}^{N} w_{r.k}^j(v)c(Y_{jk}, y), \;\; r = 2, \ldots, J$$

where

$$w_{r.k}^j(v) = \frac{K_b(v - Y_{j_rk})\mathcal{I}(\mathbf{\Delta}_k \in \widetilde{S}_r^j)}{\sum_{l=1}^{N} K_b(v - Y_{j_rl})\mathcal{I}(\mathbf{\Delta}_l \in \widetilde{S}_r^j)}, \quad \mathbf{\Delta}_k = (\Delta_{1k}, \ldots, \Delta_{ak}).$$

Note that the bandwidth can vary for different j, r combination. So the correct notation for the bandwidths used in the $w_{r.k}^j(v)$ weights is b_r^j. However, here and in the next section we suppress j and r indices and we denote all the bandwidths with b to simplify the notation.

From (3.2.4) it follows that

$$\widehat{F}_j\left(y|S_r^j\right) = \frac{1}{n_r^j}\sum_{k=1}^{N}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j)\widehat{F}_j\left(y|Y_{j_r}^0 = Y_{j_rk}, \widetilde{S}_r^j\right), \;\; r = 2, \ldots, J. \quad (3.2.5)$$

Hence, $F_j\left(y|S_r^j\right)$, $r = 2, \ldots, J$, will be estimated by averaging $\widehat{F}_j\left(y|Y_{j_r}^0 = Y_{j_rk}, \widetilde{S}_r^j\right)$ over all the cases in the sample with missingness pattern S_r^j.

Finally, plugging in the estimators in (3.2.1), we get

$$\widehat{F}_j(y) = \frac{1}{n}\sum_{k=1}^{N}\left[\Delta_{jk}c(Y_{jk}, y) + \sum_{r=2}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j)\widehat{F}_j\left(y|Y_{j_r}^0 = Y_{j_rk}, \widetilde{S}_r^j\right)\right] \quad (3.2.6)$$

for $j = 1, \ldots, J$.

Remark 3.2.2. *The MCR assumption is defined through equations of conditional cumulative distribution functions. Expressing the assumption in this form led to a smoother derivation of the nonparametric imputation. However, there is an equivalent expression of the MCR assumption using conditional probabilities, which enable easier comparison with the classical MAR assumption.*

In particular , equation (3.1.6) is equivalent to

$$P\left[(\Delta_1, \ldots, \Delta_J) = \mathbf{0} | Y_1^0, \ldots, Y_J^0\right] = P\left[(\Delta_1, \ldots, \Delta_J) = \mathbf{0}\right],$$

and the set of equations (3.2.3) is equivalent to

$$P(\Delta_j | \{S_r^j \cup \widetilde{S}_r^j\}, Y_1^0, \ldots, Y_J^0) = P(\Delta_j | \{S_r^j \cup \widetilde{S}_r^j\}, Y_{jr}^0), \quad r = 2, \ldots, J.$$

It can be seen that, perhaps the most basic difference between the two assumptions is that the set on which the MCR assumption condition contains information on the indexes of the observed coordinates as well as on the values of some Δ coordinates.

3.3 Test Statistic and Asymptotic Results

Let

$$\widehat{H}(y) = \frac{1}{n.} \sum_{j=1}^{J} \sum_{k=1}^{N} \Delta_{jk} c(Y_{jk}, y), \quad \text{where } n. = \sum_{j=1}^{J} n_j.$$

to denote the empirical distribution function obtained from all data.

Considering the null hypothesis defined in (3.1.4), the proposed test statistic is of the form

$$T_{\mathbf{C}} = \mathbf{C} \int \widehat{H}(y) d\widehat{\mathbf{F}}(y), \tag{3.3.1}$$

where, $\widehat{\mathbf{F}} = (\widehat{F}_1, \ldots, \widehat{F}_J)$, for $\widehat{F}_j$ given by (3.2.6).

More explicit,

$$T_{\mathbf{C}} = \mathbf{C}\begin{pmatrix} \int \widehat{H}(y)d\widehat{F}_1(y) \\ \vdots \\ \int \widehat{H}(y)d\widehat{F}_J(y) \end{pmatrix}$$

where,

$$\begin{aligned}\int \widehat{H}(y) \quad d \quad \widehat{F}_j(y) &= \frac{1}{n}\sum_{k=1}^{N}\Bigg[\Delta_k \widehat{H}(Y_{jk}) + \sum_{r=2}^{J}\mathcal{I}(\Delta_{jk} \in S_r^j) \\ &\quad \times \int \widehat{H}(y)d\widehat{F}_j(y|Y_{j_r}^0 = Y_{j_r k}, \widetilde{S}_r^j)\Bigg] \\ &= \frac{1}{n}\sum_{k=1}^{N}\left[\Delta_{jk}\widehat{H}(Y_{jk}) + \sum_{r=2}^{J}\mathcal{I}(\Delta_k \in S_r^j)\sum_{l=1}^{N} w_{r,l}^j(Y_{j_r k})\widehat{H}(Y_{jl})\right] \quad (3.3.2)\end{aligned}$$

Derivation of the asymptotic distribution of $T_{\mathbf{C}}$ is achieved following similar steps as deriving the one for the paired data case in Chapter 2. First note that under the null hypothesis

$$T_{\mathbf{C}} = \mathbf{C}\int \widehat{H}(y)d\left(\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right), \quad (3.3.3)$$

therefore, the main goal is to derive the asymptotic theory of the vector

$$\int \widehat{H}(y)d\left(\widehat{\mathbf{F}}(y) - \mathbf{F}(y)\right). \quad (3.3.4)$$

The approach for doing so consists of two stages. First we show that $\widehat{H}$ can be replaced by H in (3.3.4) (Proposition 3.3.1) and then we derive the asymptotic distribution of $\int H(y)d(\widehat{\mathbf{F}}(y) - \mathbf{F}(y))$ (Theorem 3.3.2 and Corollary 3.3.3).

Proposition 3.3.1. *For $H = E(\widehat{H})$ and $\widehat{F}_j$ as in (3.2.6), under Assumptions A1-A3 stated in Section 3.7, for $j = 1, \ldots, J$*

$$\sqrt{N}\int (\widehat{H} - H)d(\widehat{F}_j - F_j) \xrightarrow{p} 0, \ \text{as } N \to \infty.$$

This implies that,

$$\int (\widehat{H} - H)d(\widehat{\mathbf{F}} - \mathbf{F}) = o_p(N^{-1/2})$$

and the asymptotic distribution of $\int \widehat{H}(y)d(\widehat{\mathbf{F}}(y) - \mathbf{F}(y))$, follows from the asymptotic distribution of $\int H(y)d(\widehat{\mathbf{F}}(y) - \mathbf{F}(y))$.

Theorem 3.3.2. *Let $\widehat{F}_j$ be the estimator defined in (3.2.6), under Assumptions A1-A3 stated in Section 3.7, for each $j = 1, \ldots, J$*

$$\int H(y)d(\widehat{F}_j(y) - F_j(y)) = \frac{1}{n}\sum_{k=1}^{N} h_j(\mathbf{Z}_k) + o_p(N^{-1/2})$$

where

$$\begin{aligned} h_j(\mathbf{Z}_k) &= \left[H(Y_{jk}^0) - E\{H(Y_j^0)\}\right]\sum_{r=1}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j) \\ &\quad + \sum_{r=2}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j)\left[E\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_r k}, S_r^j\} - H(Y_{jk}^0)\right] \\ &\quad + 2n\sum_{r=2}^{J}\frac{H_{r,1}^j(\mathbf{Z}_k)}{\tilde{n}_r^j} \end{aligned}$$

for $\tilde{n}_r^j$ the number of cases out of N with missingness pattern $\tilde{S}_r^j$ and

$$\begin{aligned} H_{r,1}^j(\mathbf{Z}_k) &= E\{H_r^j(\mathbf{Z}_k, \mathbf{Z}_l)|\mathbf{Z}_k\} \\ H_r^j(\mathbf{Z}_k, \mathbf{Z}_l) &= \frac{1}{2}\left[\widetilde{H}_r^j(\mathbf{Z}_k, \mathbf{Z}_l) + \widetilde{H}_r^j(\mathbf{Z}_k, \mathbf{Z}_l)\right] \\ \widetilde{H}_r^j(\mathbf{Z}_k, \mathbf{Z}_l) &= \mathcal{I}(\mathbf{\Delta}_k \in S_r^j)\frac{K_b(Y_{j_r k} - Y_{j_r l})\mathcal{I}(\mathbf{\Delta}_l \in \widetilde{S}_r^j)}{f_{j_r}(Y_{j_r k}|\widetilde{S}_r^j)} \\ &\quad \times \left[H(Y_{jl}) - E\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_r l}, S_r^j\}\right] \end{aligned}$$

Proof. The proof of this theorem uses direct variance calculations, uniform consistency of the kernel density estimator and standard asymptotic theory for U-statistics. The proof is given in Section 3.7.

Corollary 3.3.3. *Let $h_j(\mathbf{Z}_k)$ defined as defined in Theorem 3.3.2and $\mathbf{\Sigma}$ be the $J \times J$ covariance matrix with elements $\sigma_{jj'} = \mathrm{Cov}(h_j(\mathbf{Z}_k), h_{j'}(\mathbf{Z}_k))$.*

1. *Under Assumptions A1-A3 stated in Section 3.7,*

$$\sqrt{n}\int H(y)d(\widehat{\mathbf{F}}(y) - \mathbf{F}(y)) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}).$$

2. *Let $N/n \to \lambda, \quad$ as $N \to \infty$. Then, under Assumptions A1-A3 stated in Section 3.7,*

$$\sqrt{N}\int \widehat{H}(y)d(\widehat{\mathbf{F}}(y) - \mathbf{F}(y)) \xrightarrow{d} N(\mathbf{0}, \lambda\mathbf{\Sigma}).$$

To obtain a consistent estimator of $\mathbf{\Sigma}$, we first obtain consistent estimators of $h_j(\mathbf{Z}_k),\ j = 1, \ldots, J$ and the we estimate $vSigma$ by their sample covariance matrix. Note that

$$\begin{aligned}
h_j(\mathbf{Z}_k) &= \left[H(Y_{jk}^0) - E\left(H(Y_j^0)\right)\right]\sum_{r=1}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j) \\
&\quad + \sum_{r=2}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j)\left[E\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_r k}, S_r^j\} - H(Y_{jk}^0)\right] \\
&\quad + 2n\sum_{r=2}^{J}\frac{H_{r,1}^j(\mathbf{Z}_k)}{\tilde{n}_r^j} \\
&= \mathcal{I}(\Delta_k \in S_1^j)H(Y_{jk}) + \sum_{r=2}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j)E\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_r k}, S_r^j\} \\
&\quad - E\left[H(Y_j^0)\right]\sum_{r=1}^{J}\mathcal{I}(\mathbf{\Delta}_k \in S_r^j) + 2n\sum_{r=2}^{J}\frac{H_{r,1}^j(\mathbf{Z}_k)}{\tilde{n}_r^j}
\end{aligned}$$

Consistent estimation of $h_j(\mathbf{Z}_k)$ can be done by consistently estimating each of its terms. The first term can be consistently estimated by $\mathcal{I}(\Delta_k \in S_1^j)\widehat{H}(Y_{jk})$, since Y_{jk} is observed in these cases. For the second term, note that,

$$\begin{aligned}
E\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_r k}, S_r^j\} &= \int H(y)dF_j(y|Y_{j_r}^0 = Y_{j_r k}, S_r^j) \\
&= \int H(y)dF_j(y|Y_{j_r}^0 = Y_{j_r k}, \widetilde{S}_r^j)
\end{aligned}$$

where the second equality holds under the MCR assumption. This can be consistently estimated by

$$\widehat{E}\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_rk}, S_r^j\} = \sum_{l=1}^{J} w_{r,l}^j(Y_{j_rk})\widehat{H}(Y_{jl}). \tag{3.3.5}$$

In the third term, $E\left\{H(Y_j^0)\right\}$ can be consistently estimated by $\int \widehat{H}(y)d\widehat{F}_j(y)$ (see (3.3.2)). Finally, we have a consistent estimator of $H_{r,1}^j(\mathbf{Z}_k)$

$$\widehat{H}_{r,1}^j(\mathbf{Z}_k) = \frac{1}{2n}\sum_{l=1}^{N}\left[\widehat{\widetilde{H}_r^j}(\mathbf{Z}_k, \mathbf{Z}_l) + \widehat{\widetilde{H}_r^j}(\mathbf{Z}_l, \mathbf{Z}_k)\right],$$

where

$$\begin{aligned}\widehat{\widetilde{H}_r^j}(\mathbf{Z}_u, \mathbf{Z}_v) &= \mathcal{I}(\mathbf{\Delta}_u \in S_r^j)\frac{K_b(Y_{j_ru} - Y_{j_rv})\mathcal{I}(\mathbf{\Delta}_v \in \widetilde{S}_r^j)}{\widehat{f}_{j_r}(Y_{j_ru}|\widetilde{S}_r^j)} \\ &\quad \times \left[\widehat{H}(Y_{jv}) - \widehat{E}\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_rv}, S_r^j\}\right]\end{aligned}$$

for $\widehat{f}_{j_r}(Y_{j_ru}|\widetilde{S}_r^j) = \frac{1}{\widetilde{n}_r^j}\sum_{q=1}^{N}\mathcal{I}(\mathbf{\Delta}_q \in \widetilde{S}_r^j)K_b(Y_{j_ru} - Y_{j_rq})$.

Let $\widehat{\mathbf{\Sigma}}$ to denote the sample covariance matrix of the consistent estimators of $h_j(\mathbf{Z}_k),\ j = 1, \ldots, J$.

Corollary 3.3.4. *Under Assumptions A1-A3 stated in Section 3.7, and under the null hypothesis 3.1.4, as $N \to \infty$,*

$$\frac{N}{\lambda}\left[\mathbf{C}\int \widehat{H}(y)d\widehat{\mathbf{F}}(y)\right]'\left(\mathbf{C}\widehat{\mathbf{\Sigma}}\mathbf{C}'\right)^{-1}\left[\mathbf{C}\int \widehat{H}(y)d\widehat{\mathbf{F}}(y)\right] \xrightarrow{d} \chi_{J-1}^2.$$

3.4 Extensions to higher-way factorial designs

In terms of our real data example, what we have seen so far in this chatpter is how to estimate the marginal distribution of the measure of delinquent is each of the four interviews, $(F_1, \ldots, F_4)$. However, we did not take under consideration the information given for each boy regarding prior placements or length of most recent stay at the institution. When the factors are present, we categorize each subject to one distinct factor-level combination (group), and basically estimate the marginal distribution in each group as described in Section 3.2. To describe this we need to have an additional index in our notation, in this section we consider the $\mathbf{Z}_{ik},\ k = 1, \ldots, N_i$ vectors defined in (3.1.1) to denote our data.

Under this framework, we need to estimate the marginal distribution of the response at the j-th interview of boys in the i-th group, $F_{ij},\ i = 1, \ldots, I$, $j = 1, \ldots, J$, for $J = 4$. F_{ij} is estimated using the same procedure described in Section 3.2, considering only the N_i observations in the i-th group as our sample. Letting n_i denote the number of boys in the i-th group participating in at least one interview, similar to (3.2.6) we have

$$\widehat{F}_{ij}(y) = \frac{1}{n_i}\sum_{k=1}^{N_i}\left[\Delta_{ijk}c(Y_{ijk}, y) + \sum_{r=2}^{J}\mathcal{I}(\mathbf{\Delta}_{ik} \in S_r^j)\widehat{F}_{ij}\left(y|Y_{ij_r}^0 = Y_{ij_rk}, \widetilde{S}_r^j\right)\right]$$

for $i = 1, \ldots, I\ j = 1, \ldots, J$, where

$$\widehat{F}_{ij}\left(y|Y_{ij_r}^0 = v, \widetilde{S}_r^j\right) = \sum_{k=1}^{N_i} w_{r.ik}^j(v)c(Y_{ijk}, y),\ r = 2, \ldots, J$$

with

$$w_{r.ik}^j(v) = \frac{K_{b_i}(v - Y_{ij_rk})\mathcal{I}(\mathbf{\Delta}_{ik} \in \widetilde{S}_r^j)}{\sum_{l=1}^{N_i} K_b(v - Y_{ij_rl})\mathcal{I}(\mathbf{\Delta}_{il} \in \widetilde{S}_r^j)}, \quad \mathbf{\Delta}_{ik} = (\Delta_{i1k}, \ldots, \Delta_{iak}).$$

It should be pointed out that the bandwidth can vary for different $i,\ j,\ r$ combination. The correct notation for the bandwidths used in the $w_{r.ik}^j(v)$ weights is $b_{r.i}^j$. However, we suppress j, r and i indices denoting all the bandwidths with b, as we did in the previous sections, to simplify the notation.

Note that S_r^j and $\widetilde{S}_r^j$, $r = 1, \ldots, J$ remain the same events as in the simple case with no factors since they represents case-wise missingness patterns and they depend only on j.

We define the empirical distribution function obtained from all data to be

$$\widehat{H}(y) = \frac{1}{n..}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{N_i}\Delta_{ijk}c(Y_{ijk}, y),$$

where $n.. = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$, for n_{ij} to be the number of boys in the i-th group participating in the j-th interview .

Here we might be interested in testing the hypotheses of no main effects for the time factor, and for the other factors as well as the hypotheses of no interactions among the factors. Letting $\mathbf{F} = (F_{11}, \ldots, F_{1J}, \ldots, F_{I1}, \ldots, F_{IJ})'$, these hypotheses are all of the form

$$H_0 : \mathbf{CF} = \mathbf{0}.$$

The proposed test statistic is the $(I.J \times 1)$ vector

$$T_{\mathbf{C}} = \mathbf{C}\int \widehat{H}(y)d\widehat{\mathbf{F}}(y).$$

Denoting $\mathbf{F}_i = (F_{i1}, \ldots, F_{iJ})'$, $i = 1, \ldots, I$, then $\mathbf{F} = (\mathbf{F}_1', \ldots, \mathbf{F}_I')'$ and similarly we have $\widehat{\mathbf{F}} = (\widehat{\mathbf{F}}_1', \ldots, \widehat{\mathbf{F}}_I')'$. Using this notation, the test statistic is the $(IJ \times 1)$ vector

$$T_{\mathbf{C}} = \mathbf{C}\int \widehat{H}(y)d\widehat{\mathbf{F}}(y) = \mathbf{C}\begin{pmatrix} \int \widehat{H}(y)d\widehat{\mathbf{F}}_1(y) \\ \vdots \\ \int \widehat{H}(y)d\widehat{\mathbf{F}}_b(y) \end{pmatrix}$$

The asymptotic distribution of $\int \widehat{H}(y)d(\widehat{\mathbf{F}}_i(y) - \mathbf{F}_i(y))$ can be derived using the asymptotic theory presented in the previous section. We have that

$$\sqrt{N}\int \widehat{H}(y)d(\widehat{\mathbf{F}}_i(y) - \mathbf{F}_i(y)) \xrightarrow{d} N(\mathbf{0}, \lambda_i\mathbf{\Sigma}_i),$$

where $(N/n_i) \to \lambda_i$, as $N \to \infty$, $i = 1, \ldots, I$, and $N = \sum_{i=1}^{I} N_i$.

Note that the vectors $\int \widehat{H}(y)d(\widehat{\mathbf{F}}_i(y) - \mathbf{F}_i(y)),\ i = 1, \ldots, I$, are independent, since each subject can be categorized in a distinct factor-level combination, and observations between subjects are independent. Therefore, letting $\mathbf{V}$ to be a block diagonal covariance matrix with the i-th block being $\lambda_i \boldsymbol{\Sigma}_i$, as $N \to \infty$

$$\sqrt{N} \int \widehat{H}(y)d(\widehat{\mathbf{F}}(y) - \mathbf{F}(y)) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}).$$

Let $\widehat{\boldsymbol{\Sigma}}_i$ denote the estimator of $\widehat{\boldsymbol{\Sigma}}_i$ obtained with similar steps as $\boldsymbol{\Sigma}$ described in the previous sections, and let $\widehat{\mathbf{V}}$ be the block diagonal matrix with diagonal elements $\lambda_i \widehat{\boldsymbol{\Sigma}}_i$.Then, under the null hypothesis, as $N \to \infty$

$$N \left[\mathbf{C} \int \widehat{H}(y)d\widehat{\mathbf{F}}(y) \right]' \left(\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}' \right)^{-1} \left[\mathbf{C} \int \widehat{H}(y)d\widehat{\mathbf{F}}(y) \right] \xrightarrow{d} \chi^2_\nu,$$

where $\nu = \text{rank}(\mathbf{C})$.

3.5 Simulation Results

In this section we examine the the achieved Type I error probability of the proposed test statistic under different bandwidth sizes, and investigate its power properties. We compare its performance with the normal-based linear mixed effects models proposed by Laird and Ware (1982) and the nonparametric complete-cases approach which consists of applying the statistics of Akritas and Brunner (1997) to the complete cases.

The linear mixed effects models have the form

$$\mathbf{Y}_k = \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \mathbf{b}_k + \boldsymbol{\epsilon}_k \tag{3.5.1}$$

where $\mathbf{Y}_k$ the vector of responses for subject k whose length might vary among units, $\mathbf{X}_k$ and $\mathbf{Z}_k$ are known covariate matrices, $\boldsymbol{\beta}$ is a vector of fixed effects, $\mathbf{b}_k \sim N(\mathbf{0}, \boldsymbol{\Psi})$ is a vector of random effects of subject k and $\boldsymbol{\epsilon}_k \sim N(0, \sigma^2 \mathbf{I})$ is a vector of residual errors. These models are fitted using EM-type algorithms. Although the EM iterations usually bring the parameters into the region of optimum very quickly the progress towards the optimum tends to be slow. Bates and Pinherio (1998) proposed a hybrid approach which it starts by performing a moderate number of EM iterations and then switching to Newton-Raphson iterations which close to the optimum converges very quickly. See Pinheiro and Bates (2000) for a discussion on the implementation of (3.5.1) using the nlme package in S-plus or R, but the same analysis can be carried out with the SAS PROC MIXED procedure. This model is also suitable for the analysis of missing data. As Collins, Schafer and Kam (2001) point out, ML analysis under the MAR assumption can be carried out by ignoring all missing values and feeding the observed data in, e.g., the SAS PROC MIXED procedure. Equivalent analysis can be performed using the nlme package in S-plus or R.

In our simulations we consider an one-factor design with two levels, thus we have $I = 2$ groups crossed with J time points. We consider some cases with three and some cases with four time points.

We consider three families of joint distributions to generate the vectors $\mathbf{Y}_{ik} = (Y_{i1k}, \ldots, Y_{iJk})'$, for $i = 1, 2$, $k = 1, \ldots, N_i$. The first family is a multivariate normal distribution with $E(Y_{i1k}) = 0$, $E(Y_{ijk}) = \tau(J - j + 1)^{-1}(n../2)^{-1/2}$, for $j = 2, \ldots, J$ and $Cov(Y_{ijk}, Y_{ij'k}) = \rho_i^{|j-j'|}$, for $j, j' = 1, \ldots, J$. For example, for $J = 4$,

$$\mathbf{Y}_{ik} \sim N_4 \left[\begin{pmatrix} 0 \\ \tau/(3\sqrt{n../2}) \\ \tau/(2\sqrt{n../2}) \\ \tau/\sqrt{n../2} \end{pmatrix}, \begin{pmatrix} 1 & \rho_i & \rho_i^2 & \rho_i^3 \\ \rho_i & 1 & \rho_i & \rho_i^2 \\ \rho_i^2 & \rho_i & 1 & \rho_i \\ \rho_i^3 & \rho_i^2 & \rho_i & 1 \end{pmatrix} \right].$$

The second family we consider for the vectors have marginal log-normal distributions and the data are generated as $Y_{ijk} = e^{X_{i1k}}$ where $(X_{i1k}, \ldots, X_{iJk})$ are generated from the multivariate normal distribution described above. The third family is multivariate gamma distribution such that the marginal distributions of $Y_{i1k}, \ldots, Y_{iJk}$ are Gamma, with location parameters $\mu_{i1} = 1$, and $\mu_{ij} = 1 + \tau(J - j + 1)^{-1}(n../2)^{-1/2}$, for $j = 2, \ldots, J$, scale parameters equal to one, and $Corr(Y_{ijk}, Y_{ij'k}) = \rho_i$, $i = 1, 2$, $k = 1, \ldots, N_i$. Note that $\tau = 0$ corresponds to the null hypotheses and values away from zero generate alternatives to the hypothesis of no main time main effect. Under all of the distribution families we consider cases where $\rho_1 = \rho_2 = .25$, $\rho_1 = \rho_2 = .75$, and $\rho_1 = .25$, $\rho_2 = .75$.

We consider two models to create missingness patterns satisfying. The first model is MCAR and is applied with $J = 4$. In particular $P(\Delta_{ijk} = 0) = p_{ij}$, where $p_{i1} = 0$, $p_{i2} = 0.1$, $p_{i3} = .2$, and $p_{i4} = .3$ for $i = 1, 2$. The second model is MCR and is applied with $J = 3$. In particular, we have that $P(\Delta_{i1k} = 0) = 0$ for all i and k, and

$$P(\Delta_{ijk} = 0) = \left\{1 + exp[-(\beta_0 + \beta_1 Y_{i(j-1)k})]\right\}^{-1} \Delta_{i(j-1)k} + p_{ij}(1 - \Delta_{i(j-1)k}),$$

for $j = 2, 3$, where $\beta_1 = log2$ and β_0 is such that $\left\{1 + exp[-(\beta_0 + \beta_1 Y_{i(j-1)k})]\right\}^{-1}$ evaluated at the mean value of $Y_{i(j-1)k}$ under the null hypothesis $(\tau = 0)$ is equal to p_{ij}, where $p_{11} = 0$, $p_{12} = p_{13} = .1$ and $p_{21} = 0$, $p_{22} = p_{23} = .2$.

Simulations indicated that the assumption of compact support for the kernel function is not critical for the performance of the procedure, and we chose the normal density kernel for our simulations. In all cases we consider in the simulation study the nominal level is $\alpha = .1$. The R statistical package was used and all the results are based on 1000 runs. Tables 3.1 - 3.6, report Type I error rates for the competing procedures under all null hypotheses (i.e. $\tau = 0$) for the cases where the data are MCAR, and Tables 3.7 - 3.12, for the cases where the data are MCR.

For the proposed nonparametric test we used $b_{ij} = .5$, $b_{ij} = 1$ and $b_{ij} = 2$, where b_{ij} denotes the bandwidth used for estimating the corresponding F_{ij}. Power comparisons were done only for main time effect alternatives, for $\tau = 0, 1, 2, 3, 4$. Figures 3.1 and 3.2present the results for the normal and log-normal cases, in Figures 3.1 and 3.2.

The results in the Tables 3.1 - 3.6 indicate that the complete cases analysis (CC) performs well under MCAR missingness in terms of Type I error, especially when the sample size is fairly large (for the cases where $(N_1, N_2) = (100, 100)$, $(100, 200)$, $(200, 200)$). Even though the CC test is unbiased for MCAR data, when $p_{i1} = 0$, $p_{i2} = 0.1$, $p_{i3} = .2$, and $p_{i4} = .3$ for $i = 1, 2$, approximately 50% of the cases will have at least one missing observation and therefore disregarded for this analysis. Tables 3.7- 3.12 show that the CC analysis can perform very poorly in terms of Type I error when the data are generated under MCR. sNote that the Type I error for the CC analysis for interaction is satisfactory, inspite of the biases in estimating the F_{ij}'s. This is due to the fact that, for the MCR model used in the simulations, these biases cancel out in the interaction contrasts, while they add up in the main effects contrasts.

When $\rho_1 = \rho_2$ the achieved α-level of the F-tests for main factor and interaction effects using the linear mixed effects model (FT) is satisfactory in terms of Type I error for both normal and non-normal cases. However, when $\rho_1 \neq \rho_2$ the results are liberal in all of the cases; this behavior is more severe in non-normal cases, when $N_1 \neq N_2$ or when the data is MCR.

The proposed nonparametric test (NP) has satisfactory Type I error rate in all cases, for at least one of the bandwidth values when the sample size is sufficiently large $((N_1, N_2) = (100, 100),\ (100, 200),\ (200, 200))$. When $(N_1, N_2) = (50, 50)$ and in most of the cases where $(N_1, N_2) = (50, 100)$ the results are liberal. Note that the results in the simulation study in Chapter 2 indicate that for $(N_1, N_2) = (50, 50) or (50, 100)$ the NP-method performed well in the case of two time points. As the number of time points increases, the number of possible missingness patterns increases as well, and since the NP-test requires consistent estimation of the conditional distribution functions given any missingness pattern the sample size should be larger.

In all MCAR cases the performance of the test is pretty robust to bandwidth selection since bandwidths of .5, 1 or 2 perform well for a variety of sample sizes, distribution families and correlations. In the MCR cases, however, the bandwidth selection is more crucial, especial when the data are highly correlated ($\rho_i = .75$). When $\rho_1 = \rho_2 = .75$ or $\rho_1 = .25,\ \rho_2 = .75$ bandwidth of of .5 seems to give consistently better results in all of the cases.

The power simulations, which are summarized in Figures 3.1 and 3.2 are all for the balanced case with $N_1 = N_2 = 100$. Figure 3.1 summarizes the results for MCAR normal and log-normal data for the cases $\rho_1 = \rho_2 = .25$, $\rho_1 = \rho_2 = .75$ and $\rho_1 = .25,\ \rho_2 = .75$. Similarly, Figure 3.2 summarizes the results for MCR cases. In all of this cases the power of the FT is very low compared to the power of the other tests. Figure 3.1 shows that the power of the CC test is always lower than the one of the NP-test. In the MCR cases the CC-test is highly biased and thus the power results are not of interest.

	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
Eff.	FT	CC	$NP_{.5}$	NP_1	NP_2	FT	CC	$NP_{.5}$	NP_1	NP_2
	$N_1 = 50$ $N_2 = 50$					$N_1 = 50$ $N_2 = 50$				
T	0.095	0.124	0.137	0.131	0.131	0.107	0.128	0.133	0.128	0.128
G	0.096	0.104	0.094	0.095	0.097	0.100	0.097	0.097	0.102	0.099
TG	0.102	0.128	0.146	0.134	0.137	0.104	0.127	0.147	0.142	0.133
	$N_1 = 50$ $N_2 = 100$					$N_1 = 50$ $N_2 = 100$				
T	0.105	0.116	0.129	0.118	0.120	0.112	0.120	0.116	0.111	0.117
G	0.096	0.094	0.124	0.122	0.122	0.103	0.098	0.106	0.113	0.108
TG	0.101	0.121	0.126	0.120	0.119	0.109	0.133	0.141	0.144	0.143
	$N_1 = 100$ $N_2 = 100$					$N_1 = 100$ $N_2 = 100$				
T	0.103	0.114	0.116	0.111	0.102	0.104	0.118	0.117	0.102	0.098
G	0.094	0.107	0.105	0.104	0.106	0.100	0.105	0.103	0.106	0.107
TG	0.093	0.102	0.105	0.100	0.097	0.094	0.111	0.111	0.113	0.117
	$N_1 = 100$ $N_2 = 200$					$N_1 = 100$ $N_2 = 200$				
T	0.101	0.127	0.123	0.122	0.124	0.112	0.127	0.122	0.119	0.111
G	0.089	0.109	0.115	0.112	0.108	0.097	0.102	0.105	0.113	0.106
TG	0.101	0.137	0.126	0.121	0.114	0.099	0.125	0.119	0.123	0.115
	$N_1 = 200$ $N_2 = 200$					$N_1 = 200$ $N_2 = 200$				
T	0.092	0.098	0.096	0.097	0.102	0.104	0.090	0.096	0.091	0.088
G	0.095	0.099	0.104	0.103	0.101	0.100	0.102	0.094	0.101	0.101
TG	0.094	0.116	0.109	0.105	0.105	0.102	0.110	0.103	0.095	0.097

Table 3.1: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR Normal data; Here and all tables in this Chapter T, G, TG denote the hypotheses of no main time effect, no main group effect, no interaction effect, and FT, CC, NP_b denote the F-test test, the complete pairs test, the proposed nonparametric test with bandwidths $b_{ij} = b$.

	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
Eff.	FT	CC	$NP_{.5}$	NP_1	NP_2	FT	CC	$NP_{.5}$	NP_1	NP_2
	$N_1 = 50$ $N_2 = 50$					$N_1 = 50$ $N_2 = 50$				
T	0.095	0.124	0.137	0.131	0.131	0.107	0.128	0.133	0.128	0.128
G	0.096	0.104	0.094	0.095	0.097	0.100	0.097	0.097	0.102	0.099
TG	0.102	0.128	0.146	0.134	0.137	0.104	0.127	0.147	0.142	0.133
	$N_1 = 50$ $N_2 = 100$					$N_1 = 50$ $N_2 = 100$				
T	0.105	0.116	0.129	0.118	0.120	0.112	0.120	0.116	0.111	0.117
G	0.096	0.094	0.124	0.122	0.122	0.103	0.098	0.106	0.113	0.108
TG	0.101	0.121	0.126	0.120	0.119	0.109	0.133	0.141	0.144	0.143
	$N_1 = 100$ $N_2 = 100$					$N_1 = 100$ $N_2 = 100$				
T	0.103	0.114	0.116	0.111	0.102	0.104	0.118	0.117	0.102	0.098
G	0.094	0.107	0.105	0.104	0.106	0.100	0.105	0.103	0.106	0.107
TG	0.093	0.102	0.105	0.100	0.097	0.094	0.111	0.111	0.113	0.117
	$N_1 = 100$ $N_2 = 200$					$N_1 = 100$ $N_2 = 200$				
T	0.101	0.127	0.123	0.122	0.124	0.112	0.127	0.122	0.119	0.111
G	0.089	0.109	0.115	0.112	0.108	0.097	0.102	0.105	0.113	0.106
TG	0.101	0.137	0.126	0.121	0.114	0.099	0.125	0.119	0.123	0.115
	$N_1 = 200$ $N_2 = 200$					$N_1 = 200$ $N_2 = 200$				
T	0.092	0.098	0.096	0.097	0.102	0.104	0.090	0.096	0.091	0.088
G	0.095	0.099	0.104	0.103	0.101	0.100	0.102	0.094	0.101	0.101
TG	0.094	0.116	0.109	0.105	0.105	0.102	0.110	0.103	0.095	0.097

Table 3.2: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR log-normal data.

	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
Eff.	FT	CC	$NP_{.5}$	NP_1	NP_2	FT	CC	$NP_{.5}$	NP_1	NP_2
	$N_1 = 50$ $N_2 = 50$					$N_1 = 50$ $N_2 = 50$				
T	0.109	0.125	0.145	0.142	0.140	0.105	0.159	0.130	0.126	0.135
G	0.090	0.104	0.112	0.106	0.103	0.071	0.084	0.077	0.078	0.075
TG	0.097	0.124	0.125	0.126	0.121	0.100	0.128	0.116	0.108	0.110
	$N_1 = 50$ $N_2 = 100$					$N_1 = 50$ $N_2 = 100$				
T	0.109	0.124	0.132	0.125	0.119	0.114	0.119	0.131	0.132	0.126
G	0.101	0.120	0.122	0.121	0.121	0.096	0.119	0.123	0.114	0.113
TG	0.107	0.133	0.116	0.117	0.115	0.110	0.136	0.126	0.123	0.119
	$N_1 = 100$ $N_2 = 100$					$N_1 = 100$ $N_2 = 100$				
T	0.109	0.102	0.111	0.108	0.107	0.104	0.127	0.136	0.138	0.129
G	0.118	0.082	0.106	0.100	0.096	0.112	0.098	0.117	0.120	0.114
TG	0.105	0.118	0.112	0.113	0.112	0.099	0.109	0.112	0.102	0.096
	$N_1 = 100$ $N_2 = 200$					$N_1 = 100$ $N_2 = 200$				
T	0.091	0.122	0.137	0.129	0.134	0.083	0.125	0.110	0.113	0.097
G	0.105	0.104	0.094	0.096	0.095	0.080	0.104	0.108	0.107	0.109
TG	0.088	0.118	0.081	0.085	0.086	0.090	0.103	0.100	0.091	0.087
	$N_1 = 200$ $N_2 = 200$					$N_1 = 200$ $N_2 = 200$				
T	0.097	0.108	0.096	0.099	0.097	0.100	0.110	0.096	0.101	0.101
G	0.080	0.101	0.101	0.102	0.104	0.089	0.098	0.095	0.099	0.103
TG	0.095	0.097	0.077	0.075	0.077	0.117	0.104	0.116	0.120	0.128

Table 3.3: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR gamma data.

$\rho_1 = .25,\ \rho_2 = .75$	Effect	FT	CC	$NP_{.5}$	NP_1	NP_2
$N_1 = 50,\ N_2 = 50$	T	0.128	0.130	0.120	0.113	0.109
	G	0.100	0.107	0.107	0.111	0.109
	TG	0.097	0.132	0.150	0.140	0.135
$N_1 = 50,\ N_2 = 100$	T	0.143	0.128	0.141	0.133	0.132
	G	0.112	0.092	0.114	0.117	0.113
	TG	0.133	0.138	0.141	0.141	0.144
$N_1 = 100,\ N_2 = 100$	T	0.137	0.132	0.121	0.116	0.110
	G	0.103	0.112	0.110	0.112	0.117
	TG	0.099	0.123	0.119	0.112	0.117
$N_1 = 100,\ N_2 = 200$	T	0.160	0.119	0.122	0.115	0.111
	G	0.104	0.120	0.107	0.104	0.102
	TG	0.140	0.138	0.130	0.125	0.126
$N_1 = 200,\ N_2 = 200$	T	0.114	0.105	0.107	0.099	0.103
	G	0.097	0.093	0.099	0.098	0.095
	TG	0.105	0.107	0.102	0.098	0.091

Table 3.4: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR normal data; $\rho_1 = .25$ and $\rho_2 = .75$.

$\rho_1 = .25,\ \rho_2 = .75$	Effect	FT	CC	$NP_{.5}$	NP_1	NP_2
$N_1 = 50,\ N_2 = 50$	T	0.121	0.131	0.160	0.148	0.137
	G	0.090	0.099	0.108	0.109	0.109
	TG	0.112	0.146	0.169	0.155	0.155
$N_1 = 50,\ N_2 = 100$	T	0.138	0.138	0.122	0.119	0.116
	G	0.093	0.102	0.118	0.113	0.111
	TG	0.125	0.130	0.144	0.135	0.132
$N_1 = 100,\ N_2 = 100$	T	0.130	0.100	0.115	0.103	0.105
	G	0.093	0.103	0.107	0.103	0.100
	TG	0.105	0.126	0.118	0.109	0.105
$N_1 = 100,\ N_2 = 200$	T	0.143	0.112	0.117	0.112	0.115
	G	0.114	0.113	0.125	0.126	0.121
	TG	0.118	0.145	0.136	0.137	0.132
$N_1 = 200,\ N_2 = 200$	T	0.123	0.114	0.102	0.101	0.097
	G	0.113	0.099	0.099	0.102	0.099
	TG	0.106	0.111	0.125	0.125	0.121

Table 3.5: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR log-normal data; $\rho_1 = .25$ and $\rho_2 = .75$.

$\rho_1 = .25,\ \rho_2 = .75$	Effect	FT	CC	$NP_{.5}$	NP_1	NP_2
$N_1 = 50,\ N_2 = 50$	T	0.138	0.119	0.141	0.133	0.132
	G	0.095	0.096	0.109	0.108	0.109
	TG	0.101	0.120	0.118	0.113	0.115
$N_1 = 50,\ N_2 = 100$	T	0.218	0.151	0.162	0.156	0.148
	G	0.113	0.095	0.109	0.099	0.100
	TG	0.188	0.146	0.139	0.134	0.133
$N_1 = 100,\ N_2 = 100$	T	0.134	0.129	0.107	0.104	0.104
	G	0.093	0.087	0.096	0.093	0.098
	TG	0.090	0.115	0.106	0.096	0.099
$N_1 = 100,\ N_2 = 200$	T	0.200	0.130	0.113	0.112	0.112
	G	0.119	0.096	0.121	0.124	0.124
	TG	0.168	0.114	0.110	0.110	0.110
$N_1 = 200,\ N_2 = 200$	T	0.150	0.103	0.100	0.111	0.117
	G	0.103	0.106	0.116	0.119	0.116
	TG	0.095	0.114	0.098	0.096	0.091

Table 3.6: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR gamma data; $\rho_1 = .25$ and $\rho_2 = .75$.

	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
Eff.	FT	CC	$NP_{.5}$	NP_1	NP_2	FT	CC	$NP_{.5}$	NP_1	NP_2
	$N_1 = 50$ $N_2 = 50$					$N_1 = 50$ $N_2 = 50$				
T	0.103	0.177	0.112	0.109	0.110	0.115	0.137	0.131	0.145	0.183
G	0.098	0.147	0.120	0.117	0.117	0.114	0.158	0.109	0.111	0.112
TG	0.110	0.130	0.112	0.108	0.111	0.105	0.113	0.118	0.112	0.11
	$N_1 = 50$ $N_2 = 100$					$N_1 = 50$ $N_2 = 100$				
T	0.094	0.198	0.109	0.112	0.112	0.100	0.148	0.119	0.160	0.217
G	0.094	0.143	0.092	0.100	0.104	0.093	0.163	0.101	0.102	0.110
TG	0.087	0.120	0.100	0.097	0.099	0.091	0.113	0.112	0.104	0.100
	$N_1 = 100$ $N_2 = 100$					$N_1 = 100$ $N_2 = 100$				
T	0.106	0.235	0.110	0.111	0.122	0.111	0.171	0.109	0.153	0.240
G	0.091	0.179	0.119	0.122	0.124	0.095	0.191	0.104	0.111	0.115
TG	0.109	0.117	0.101	0.099	0.101	0.120	0.121	0.119	0.119	0.122
	$N_1 = 100$ $N_2 = 200$					$N_1 = 100$ $N_2 = 200$				
T	0.087	0.272	0.093	0.091	0.102	0.094	0.180	0.106	0.156	0.269
G	0.089	0.172	0.106	0.105	0.109	0.082	0.226	0.103	0.110	0.109
TG	0.087	0.124	0.099	0.093	0.096	0.087	0.120	0.106	0.117	0.118
	$N_1 = 200$ $N_2 = 200$					$N_1 = 200$ $N_2 = 200$				
T	0.098	0.349	0.104	0.105	0.115	0.098	0.224	0.127	0.217	0.384
G	0.100	0.202	0.100	0.099	0.104	0.097	0.254	0.098	0.112	0.113
TG	0.090	0.111	0.088	0.087	0.088	0.087	0.109	0.094	0.097	0.114

Table 3.7: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR normal data.

	$\rho_1 = \rho_2 = .25$					$\rho_1 = \rho_2 = .75$				
Eff.	FT	CC	$NP_{.5}$	NP_1	NP_2	FT	CC	$NP_{.5}$	NP_1	NP_2
	$N_1 = 50 \;\; N_2 = 50$					$N_1 = 50 \;\; N_2 = 50$				
T	0.105	0.273	0.148	0.117	0.112	0.104	0.166	0.132	0.126	0.197
G	0.101	0.143	0.137	0.124	0.125	0.105	0.153	0.130	0.127	0.136
TG	0.109	0.140	0.127	0.109	0.103	0.118	0.150	0.141	0.112	0.109
	$N_1 = 50 \;\; N_2 = 100$					$N_1 = 50 \;\; N_2 = 100$				
T	0.093	0.295	0.141	0.127	0.122	0.106	0.177	0.137	0.149	0.241
G	0.101	0.149	0.134	0.123	0.120	0.098	0.172	0.117	0.116	0.121
TG	0.103	0.108	0.128	0.111	0.103	0.097	0.104	0.128	0.118	0.106
	$N_1 = 100 \;\; N_2 = 100$					$N_1 = 100 \;\; N_2 = 100$				
T	0.108	0.382	0.141	0.128	0.124	0.125	0.203	0.130	0.153	0.302
G	0.092	0.159	0.111	0.106	0.107	0.102	0.166	0.111	0.108	0.108
TG	0.087	0.105	0.119	0.106	0.093	0.116	0.107	0.116	0.102	0.097
	$N_1 = 100 \;\; N_2 = 200$					$N_1 = 100 \;\; N_2 = 200$				
T	0.095	0.473	0.119	0.105	0.100	0.115	0.245	0.116	0.148	0.348
G	0.111	0.176	0.118	0.110	0.106	0.101	0.191	0.115	0.110	0.114
TG	0.104	0.126	0.139	0.121	0.112	0.105	0.116	0.122	0.113	0.116
	$N_1 = 200 \;\; N_2 = 200$					$N_1 = 200 \;\; N_2 = 200$				
T	0.097	0.612	0.130	0.104	0.105	0.094	0.300	0.121	0.175	0.444
G	0.099	0.195	0.119	0.110	0.111	0.098	0.229	0.112	0.110	0.120
TG	0.087	0.127	0.147	0.124	0.112	0.100	0.118	0.119	0.116	0.129

Table 3.8: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR log-normal data.

	$\rho_1=\rho_2=.25$					$\rho_1=\rho_2=.75$				
Eff.	FT	CC	$NP_{.5}$	NP_1	NP_2	FT	CC	$NP_{.5}$	NP_1	NP_2
	$N_1=50$ $N_2=50$					$N_1=50$ $N_2=50$				
T	0.110	0.165	0.127	0.115	0.114	0.107	0.143	0.125	0.147	0.191
G	0.107	0.123	0.109	0.105	0.099	0.092	0.144	0.090	0.108	0.112
TG	0.103	0.132	0.132	0.124	0.118	0.113	0.109	0.117	0.117	0.124
	$N_1=50$ $N_2=100$					$N_1=50$ $N_2=100$				
T	0.090	0.172	0.111	0.108	0.107	0.110	0.130	0.115	0.141	0.216
G	0.090	0.120	0.103	0.100	0.093	0.099	0.138	0.096	0.097	0.096
TG	0.080	0.118	0.115	0.109	0.106	0.111	0.116	0.129	0.116	0.121
	$N_1=100$ $N_2=100$					$N_1=100$ $N_2=100$				
T	0.101	0.204	0.106	0.098	0.095	0.086	0.150	0.122	0.164	0.264
G	0.094	0.135	0.108	0.106	0.102	0.094	0.150	0.096	0.094	0.100
TG	0.110	0.119	0.115	0.112	0.104	0.097	0.120	0.118	0.108	0.109
	$N_1=100$ $N_2=200$					$N_1=100$ $N_2=200$				
T	0.114	0.232	0.128	0.120	0.121	0.096	0.142	0.127	0.189	0.310
G	0.110	0.146	0.101	0.094	0.095	0.098	0.161	0.103	0.111	0.108
TG	0.109	0.139	0.121	0.113	0.111	0.088	0.117	0.116	0.117	0.129
	$N_1=200$ $N_2=200$					$N_1=200$ $N_2=200$				
T	0.099	0.278	0.114	0.113	0.110	0.096	0.134	0.123	0.222	0.439
G	0.080	0.173	0.108	0.104	0.104	0.103	0.218	0.102	0.109	0.107
TG	0.089	0.134	0.112	0.106	0.099	0.110	0.106	0.102	0.112	0.115

Table 3.9: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR gamma data.

$\rho_1 = .25,\ \rho_2 = .75$	Effect	FT	CC	$NP_{.5}$	NP_1	NP_2
$N_1 = 50,\ N_2 = 50$	T	0.142	0.150	0.124	0.138	0.148
	G	0.103	0.218	0.110	0.117	0.121
	TG	0.105	0.120	0.122	0.114	0.119
$N_1 = 50,\ N_2 = 100$	T	0.165	0.177	0.113	0.122	0.139
	G	0.113	0.259	0.106	0.112	0.124
	TG	0.140	0.121	0.102	0.108	0.103
$N_1 = 100,\ N_2 = 100$	T	0.147	0.188	0.116	0.129	0.164
	G	0.087	0.304	0.104	0.120	0.140
	TG	0.122	0.111	0.107	0.111	0.126
$N_1 = 100,\ N_2 = 200$	T	0.162	0.200	0.095	0.118	0.152
	G	0.112	0.409	0.115	0.131	0.143
	TG	0.141	0.112	0.119	0.119	0.142
$N_1 = 200,\ N_2 = 200$	T	0.146	0.249	0.105	0.148	0.217
	G	0.101	0.451	0.102	0.118	0.138
	TG	0.115	0.097	0.086	0.103	0.134

Table 3.10: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCAR normal data; $\rho_1 = .25$ and $\rho_2 = .75$.

$\rho_1 = .25,\ \rho_2 = .75$	Effect	FT	CC	$NP_{.5}$	NP_1	NP_2
$N_1 = 50,\ N_2 = 50$	T	0.149	0.202	0.131	0.120	0.127
	G	0.099	0.241	0.141	0.134	0.139
	TG	0.122	0.123	0.114	0.113	0.120
$N_1 = 50,\ N_2 = 100$	T	0.179	0.204	0.139	0.131	0.145
	G	0.122	0.278	0.130	0.123	0.132
	TG	0.150	0.116	0.122	0.112	0.114
$N_1 = 100,\ N_2 = 100$	T	0.176	0.273	0.127	0.122	0.174
	G	0.112	0.345	0.112	0.108	0.119
	TG	0.132	0.124	0.137	0.132	0.127
$N_1 = 100,\ N_2 = 200$	T	0.235	0.299	0.112	0.121	0.178
	G	0.128	0.449	0.106	0.115	0.136
	TG	0.213	0.118	0.137	0.128	0.138
$N_1 = 200,\ N_2 = 200$	T	0.208	0.409	0.120	0.128	0.207
	G	0.114	0.538	0.118	0.129	0.171
	TG	0.165	0.128	0.135	0.135	0.172

Table 3.11: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR log-normal data; $\rho_1 = .25$ and $\rho_2 = .75$.

$\rho_1 = .25,\ \rho_2 = .75$	Effect	FT	CC	$NP_{.5}$	NP_1	NP_2
$N_1 = 50,\ N_2 = 50$	T	0.145	0.125	0.125	0.132	0.149
	G	0.106	0.176	0.106	0.112	0.113
	TG	0.111	0.116	0.131	0.135	0.143
$N_1 = 50,\ N_2 = 100$	T	0.221	0.145	0.101	0.101	0.123
	G	0.122	0.208	0.108	0.115	0.128
	TG	0.181	0.127	0.098	0.100	0.108
$N_1 = 100,\ N_2 = 100$	T	0.149	0.148	0.118	0.129	0.158
	G	0.101	0.253	0.106	0.116	0.126
	TG	0.093	0.119	0.105	0.106	0.121
$N_1 = 100,\ N_2 = 200$	T	0.212	0.137	0.105	0.116	0.167
	G	0.127	0.305	0.107	0.115	0.133
	TG	0.179	0.105	0.096	0.111	0.128
$N_1 = 200,\ N_2 = 200$	T	0.156	0.172	0.106	0.141	0.202
	G	0.089	0.369	0.109	0.115	0.143
	TG	0.100	0.118	0.103	0.122	0.150

Table 3.12: Type I error rates at nominal $\alpha = 0.1$ in simulations with MCR gamma data; $\rho_1 = .25$ and $\rho_2 = .75$.

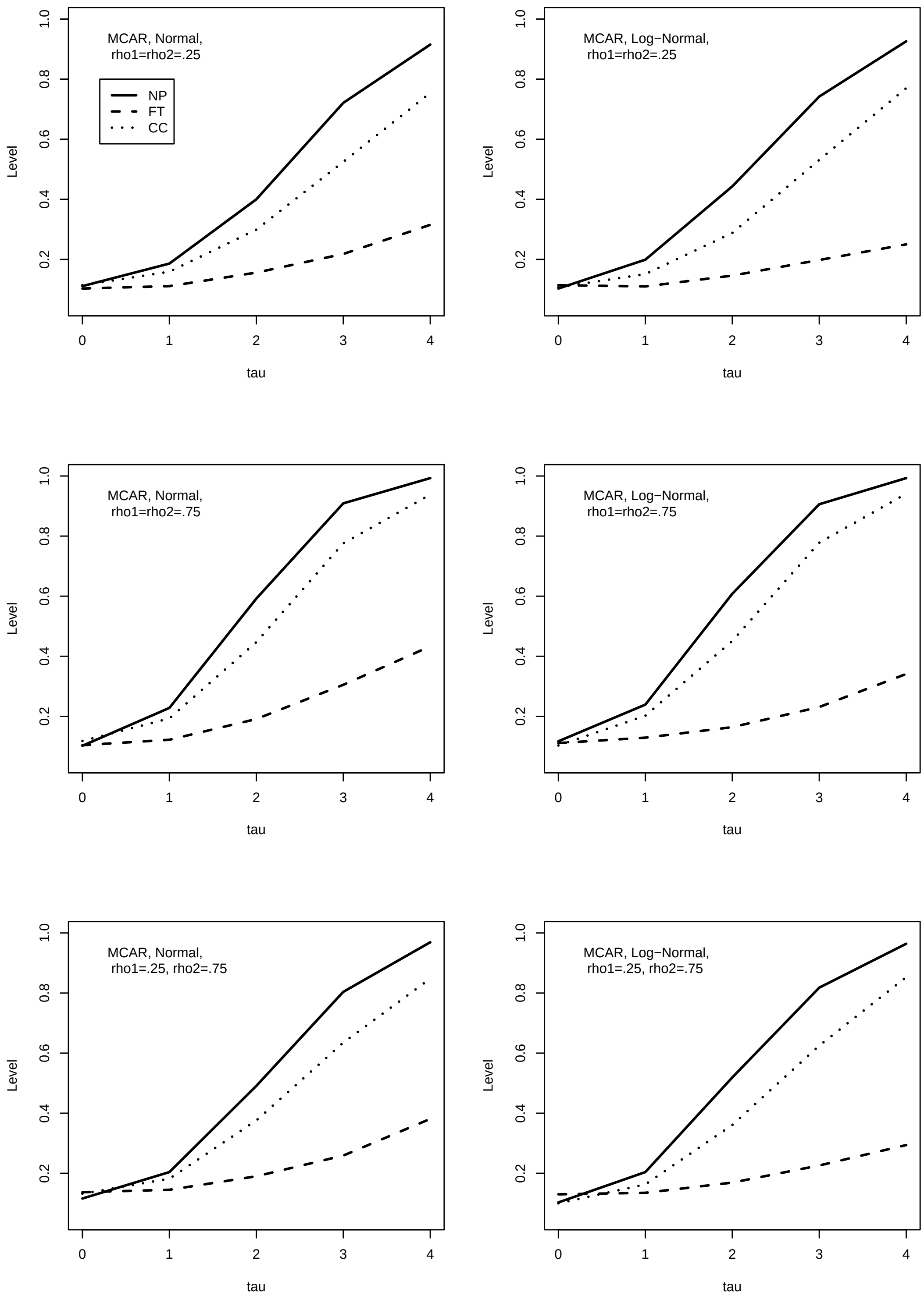

Figure 3.1: MCAR cases, $N_1 = N_2 = 100$ and for NP method, $b = 1$.

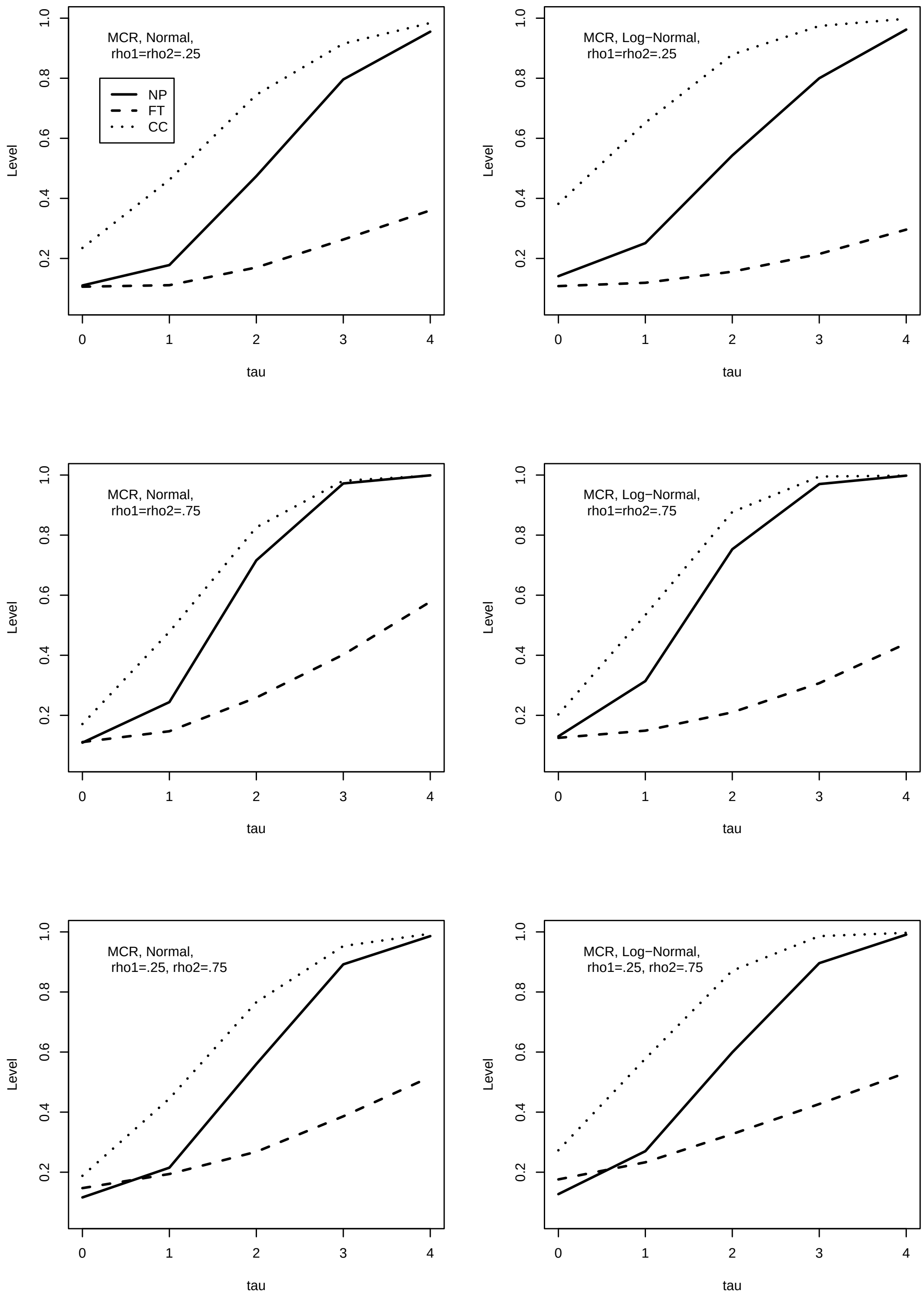

Figure 3.2: MCR cases, $N_1 = N_2 = 100$ and for NP method, $b = .5$.

3.6 Data Analysis

As an application we will consider the study of incarcerated boys in Michigan juvenile correction institutions that is described in Section ??. Recall that the interviews took place within 10 days of their arrival, after being there four months, just before leaving, six months after leaving and 36 months after leaving. During the institutional stay the response rates remained high, with 99%, 96%, 91% for the first three interviews, and 75%, 60% for the later two interviews after the boys returned to the community.

Our analysis focuses on analyzing a measure of delinquent values obtained in the first four interviews considering three binary factors type of institution (state or private), number of prior placements (0, > 0) and race (African American or not).

In our analysis we considered four tests: the F-test of the normal based linear mixed effect model, the nonparametric complete-cases test and our proposed test with two different bandwidths, and normal kernel. We also considered the Epanechikov kernel but the results were very similar. The bandwidth values we used are $b_i = \nu s_p$, $i = 1, \ldots, 8$, where s_p denotes the pooled, across the factor level combinations, standard deviation of the observed data and $\nu = .5$ and 1. The analysis for the F-test was performed using the nmle package in R which analyzes linear mixed effects according to Laird and Ware (1982). Table 3.13 gives p-values for the tests of main effects and interactions for all analyses.

The tests appear in agreement regarding the significance or not of most main effects and interactions. Discrepancies appear in the main for the number of previous placements and, perhaps more interesting from the scientific point of view, in the race-institution interaction effect. Given the discreteness of the

Effects	FT	CC	$NP_{.5}$	NP_1
Institution (I)	0.429	0.930	0.766	0.817
Afr. American (A)	0.546	0.236	0.167	0.179
Prev. Placements (P)	0.001	0.056	0.013	0.015
Time (T)	$< .001$	$< .001$	$< .001$	$< .001$
(IA)	0.150	0.074	0.033	0.031
(IP)	0.293	0.618	0.774	0.796
(IT)	$< .001$	0.006	0.006	0.004
(AP)	0.925	0.568	0.530	0.535
(AT)	0.380	0.896	0.921	0.919
(PT)	0.015	0.743	0.350	0.305
(IAP)	0.453	0.265	0.198	0.191
(IAT)	0.571	0.930	0.992	0.993
(IPT)	0.723	0.171	0.091	0.087
(APT)	0.143	0.455	0.475	0.480
(IAPT)	0.422	0.327	0.157	0.162

Table 3.13: p-values for main effects and interactions using different methods. FT, CC, and NP_ν denote the linear mixed effects F-test, the complete cases test and the proposed nonparametric test with $b_i = \nu s_p$, $i = 1, \ldots, 8$. (See text for details.)

data and the probable inappropriateness of the MCAR assumption, the significance (p-value$\simeq$ 0.03) indicated by our nonparametric procedures appears more credible than that indicated by the other two procedures.

3.7 Outline of the Proofs

The proofs of the results presented in Section 3.3 will be presented under the following technical assumptions.

Assumption A1 (i) If Y_j^0 is continuous, the corresponding bandwidth sequence b_r^j, satisfies $N(b_r^j)^4 \to 0$ and $N(b_r^j)^{3+2\delta}(\log N)^{-1} \to \infty$, as $N \to \infty$, for some $\delta > 0$.

(iii) If Y_{ij}^0 is discrete, (b_r^j) is only required to converge to zero.

Assumption A2 (i) The probability density function K is symmetric and has compact support.

(ii) The probability density function K has bounded second derivative and satisfies $\int uK(u)du = 0$.

Assumption A3 (i) If Y_j^0 is discrete, then $f_j(y|\widetilde{S}_r^j)$ of Y_j^0 given $\widetilde{S}_r^j$ satisfies $min\{f_j(y|\widetilde{S}_r^j); y \in C_j)\} > 0$, for all r and j, where C_j is the set of mass points or atoms of the distribution of Y_j^0.

(ii) If Y_j^0 is continuous, then for all j,

(a) The support $\mathcal{S}_j$ of Y_j^0 is bounded for all j;

(b) the conditional density $f_j(y|\widetilde{S}_r^j)$ satisfies $\inf\{f_j(y|\widetilde{S}_r^j); y \in \mathcal{S}_j\} > 0$, for all j and r;

(c) The first two derivatives of $F_j(y|F_{j_r}^{-1}(u|S_r^j), Y_{j_r}^0 = Y_{j_r k}, S_r^j)$ with respect to u exist and are bounded for all u, y, r and j.

(d) Let $g_{j.r}(y) = f_{j_r}\{y|S_r^j)$. The first two derivatives of $g_{j.r}(y)$ exist and are bounded for all r and j.

Proof of Proposition 3.3.1. For notation simplification, let $\mathcal{I}(S_{r.k}^j) = \mathcal{I}(\mathbf{\Delta}_k \in S_r^j)$ and $\mathcal{I}(\widetilde{S}_{r.k}^j) = \mathcal{I}(\mathbf{\Delta}_k \in \widetilde{S}_r^j)$. Using (3.2.5) and (3.2.6), we have the following decomposition of $\widehat{F}_j - F_j$

$$\begin{aligned}
\widehat{F}_j(y) \quad - \quad F_j(y) &= \frac{1}{n}\sum_{k=1}^{N}\left[\Delta_{jk}c(Y_{jk},y) + \sum_{r=2}^{J} I(S^j_{r.k})\widehat{F}_j(y|Y^0_{j_r} = Y_{j_rk}, \widetilde{S}^j_r)\right] - F_j(y) \\
&= \frac{1}{n}\sum_{k=1}^{N}[c(Y_{jk},y) - F_j(y)]\sum_{r=1}^{J} I(S^j_{r.k}) \\
&+ \frac{1}{n}\sum_{r=2}^{J}\sum_{k=1}^{N} I(S^j_{r.k})\left[F_j(y|Y^0_{j_r} = Y_{j_rk}, \widetilde{S}^j_r) - c(Y^0_{jk},y)\right] \\
&+ \frac{1}{n}\sum_{r=2}^{J}\sum_{k=1}^{N} I(S^j_{r.k})\left[\sum_{l=1}^{N} w^j_{r.l}(Y_{j_rk})\left\{c(Y_{jl},y) - F_j(y|Y^0_{j_r} = Y_{j_rl}, \widetilde{S}^j_r)\right\}\right] \\
&+ \frac{1}{n}\sum_{r=2}^{J}\sum_{k=1}^{N} I(S^j_{r.k})\left[\sum_{l=1}^{N} w^j_{r.l}(Y_{j_rk})\left\{F_j(y|Y^0_{j_r} = Y_{j_rl}, \widetilde{S}^j_r)\right.\right. \\
&\qquad \left.\left. - F_j(y|Y^0_{j_r} = Y_{j_rk}, \widetilde{S}^j_r)\right\}\right]
\end{aligned}$$

Note that the first term on the right hand side is centered, since using the first part of the MCR assumption,(3.1.6), $(Y_{j1},\ldots,Y_{jJ})$ and $\sum_{r=1}^{J} I(S^j_{r.k})$ are independent. The second and the third term on the right hand side are centered conditionally on Y_{j_rk}'s, and thus also unconditionally. The final term is the bias term.

Using the integration by parts formula, we have

$$\int(\widehat{H} - H)d(\widehat{F}_j - F_j) = -\int(\widehat{F}_j - F_j)d(\widehat{H} - H)$$

Thus in order to show that $N^{1/2}\int(\widehat{H} - H)d(\widehat{F}_j - F_j) \to 0$ it suffices to show that

$$N^{1/2}\int T_q d(\widehat{H} - H) \to 0, \quad \text{for } q = 1,\ldots,4, \tag{3.7.1}$$

where T_1, T_2, T_3 and T_4 are the first, second third and fourth term on the right hand side of the decomposition of $\widehat{F}_j - F_j$.

The proof uses similar arguments as the proof of Proposition 2.3.1. and thus is omitted. The only difference is the generalization of the notation to accommodate the different missingness patterns

Proof of Theorem 3.3.2. Using the decomposition of $\widehat{F}_j - F_j$ given in the proof of Proposition (3.3.1) we have that

$$\begin{aligned}
\int H(y) \quad d \quad (\widehat{F}_j - F_j) &= \frac{1}{n}\sum_{k=1}^{N}\left[H(Y_{jk}^0) - E\{H(Y_j^0)\}\right]\sum_{r=1}^{J}\mathcal{I}(S_{r.k}^j) \\
&+ \quad \frac{1}{n}\sum_{k=1}^{N}\sum_{r=2}^{J}\mathcal{I}(S_{r.k}^j)\left[E\{H(Y_j^0)|Y_{j_r}^0 = Y_{j_rk}, S_r^j\} - H(Y_{jk}^0)\right] \\
&+ \quad \frac{1}{n}\sum_{k=1}^{N}\sum_{r=2}^{J}I(S_{r.k}^j)\left[\sum_{s=1}^{N}w_{r.s}^j(Y_{j_rk})\{H(Y_{js}) - E(H(Y_j^0)|Y_{j_r}^0 = Y_{j_rs}, \widetilde{S}_r^j)\}\right] \\
&+ \quad \frac{1}{n}\sum_{k=1}^{N}\sum_{r=2}^{J}I(S_{r.k}^j)\sum_{s=1}^{N}w_{r.s}^j(Y_{j_rk})\int H(y)d\left\{F_j(y|Y_{j_r}^0 = Y_{j_rs}, S_r^j)\right. \\
&\left. - \; F_j(y|Y_{j_r}^0 = Y_{j_rk}, S_r^j)\right\} + o_p(N^{-1/2}).
\end{aligned}$$

therefore it suffices to show that

$$\begin{aligned}
\frac{1}{n}\sum_{r=2}^{J}\sum_{k=1}^{N}\mathcal{I}(S_{r.k}^j)&\left[\sum_{s=1}^{N}w_{r.s}^j(Y_{j_rk})\{H(Y_{js}) - E(H(Y_j^0)|Y_{j_r}^0 = Y_{j_rs}, \widetilde{S}_r^j)\}\right] \\
&= 2\sum_{k=1}^{N}\sum_{r=2}^{J}\frac{H_{r,1}^j(\mathbf{Z}_k)}{\tilde{n}_r^j} + o_p(N^{-1/2})
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n}\sum_{r=2}^{J}\sum_{k=1}^{N}I(S_{r.k}^j)\sum_{s=1}^{N}w_{r.s}^j(Y_{j_rk})&\int H(y)d\{F_j(y|Y_{j_r}^0 = Y_{j_rs}, S_r^j) \\
&- F_j(y|Y_{j_r}^0 = Y_{j_rk}, S_r^j)\} = o_p(N^{-1/2}).
\end{aligned}$$

The proof follows using similar steps as in the proofs of (2.6.9) and (2.6.10).

Chapter 4

Fully Nonparametric ANCOVA with Fixed Window Sizes

4.1 Introduction

The method of analysis of covariance is among the most commonly used statistical procedures in scientific investigation. The classical analysis of covariance model imposes quite restrictive assumptions, such as the response variable has normal distribution, the conditional variances are constant, and the conditional mean is a linear function of the covariate. These assumptions are not always valid in practice. In response to these restrictions, Akritas, Arnold and Du (2000) proposed the following fully nonparametric model for nonlinear analysis of covariance. Suppose we observe (Y_{ij}, X_{ij}), $i = 1, \ldots, a$, j=1,...,n_i, where Y_{ij} and X_{ij} represent the response and the covariate in the jth observation of the ith group. With this notation, the fully nonparametric model assumes only that

$$Y_{ij}|X_{ij} = x \sim F_{ix} \ , \tag{4.1.1}$$

where F_{ix} is defined in (1.2.2). Note that model (4.1.1) does not specify how the response distribution changes when the levels, or covariate value changes, and does not assume continuity of the conditional distributions. Thus it is completely nonparametric (also nonlinear and non-additive). In order to define effects and hypotheses in this nonparametric context, we choose a distribution function $G(x)$ and set

$$\overline{F}_{i\cdot}(y) = \int F_{ix}(y)\, dG(x), \text{ and } \overline{F}_{\cdot x}(y) = \frac{1}{a}\sum_{i=1}^{a} F_{ix}(y) \ . \tag{4.1.2}$$

If the X_{ij} are fixed, or for an analysis conditional on the observed covariate values, Akritas, Arnold and Du (2000) recommend that G be taken as $\widehat{G}(x) = N^{-1}\sum_{i=1}^{a}\sum_{j=1}^{n_i} I(X_{ij} \le x)$, where $N = \sum_{i=1}^{a} n_i$, and $I(A)$ denotes the indicator function of the event A, while if the X_{ij} are random, they recommend $G(x) = E\left[\widehat{G}(x)\right]$. Note that if the covariate has the same distribution in all groups, $\overline{F}_{i\cdot}(y)$ is the marginal distribution function of Y_{ij}.

Akritas, Arnold and Du (2000) considered weighted (mid-)rank procedures for testing hypotheses of interest regarding the $\overline{F}_{i\cdot}$. For example, if i enumerates the levels of only one factor it is of interest to test that $\overline{F}_{i\cdot}$ does not depend on i (no covariate-adjusted main factor effect), but their formulation also includes testing for no covariate-adjusted interactions when i enumerates the levels of two or more factors; see Akritas, Arnold and Du (2000) for details.

Their approach relies on consistent estimation (using kernel methods) of the conditional distribution functions F_{ix}. By its nature, application of this approach requires determination of the window bandwidth which is particularly cumbersome in this context. Here we consider an alternative test procedure for the same hypotheses. The proposed procedure does not require consistent estimation of the F_{ix}. and in particular, is using nearest neighbor windows of *fixed* size. Simulations show that window sizes of 5 and 7 perform well in a variety of situations. This is of great importance for the practical applicability of the fully nonparametric methodology. At the theoretical level, the asymptotic theory uses many elegant and novel arguments.

We also remark that Tsangari and Akritas (2003) generalized the methodology of Akritas, Arnold and Du (2000) to include ANCOVA designs with two and three covariates. However, curse of dimensionality effects prevent further generalization to more covariates. Since the methodology presented here uses fixed window sizes, it will be less affected by the curse of dimensionality and thus holds the promise of extendability to more than three covariates. Also the proposed ANCOVA procedure is dealing with cases without missing observations. In Chapter 5 we discuss a some ideas about how to extend this methodology to cases with more than one covariate and with missing data.

The basic idea underlying the proposed methodology is to think of the covariate as another factor, and consider test statistics used in ANOVA when one of the factors has many levels; see Wang and Akritas (2002). The procedures used in ANOVA with many factor levels are not directly applicable because they require at least two observations per factor level combination. The way this difficulty is overcome is described in the next section. The asymptotic distribution of the test statistic is derived under both the null and the alternative hypotheses, with the covariate being either random or non-random. In order to keep the arguments simple, we consider test statistics based on the original observations rather than the ranks.

A different perspective of the one-way design, which allows easy generalization to multi-way designs, is obtained via the unique decomposition:

$$F_{ix}(y) = M(y) + A_i(y) + D_x(y) + C_{ix}(y), \tag{4.1.3}$$

where $M(y) = a^{-1}\sum_{i=1}^{a}\int F_{ix}(y)dG(x)$, $A_i(y) = \overline{F}_{i\cdot}(y) - M(y)$, $D_x(y) = \overline{F}_{\cdot x}(y) - M(y)$ and $C_{ix}(y) = F_{ix}(y) - M(y) - A_i(y) - D_x(y)$. In this decomposition, A_i are the covariate-adjusted nonparametric main effects of the factor,

D_x is the nonparametric covariate effect and C_{ix} the nonparametric interaction between the factor and the covariate. The aforementioned hypothesis of no main factor effect is equivalent to $A_i = 0$, for all i. It is important to keep in mind that this hypothesis (as well as any other hypothesis involving the $\overline{F}_{i\cdot}$, such as no covariate adjusted main effects and interactions in multi-factor ANCOVA) can be represented in the form

$$H_0(\mathbf{C}): \ \mathbf{CA} = \mathbf{0}, \quad \text{where } \mathbf{A} \text{ is the vector of all } A_{(i)}, \tag{4.1.4}$$

and $\mathbf{C}$ is a contrast matrix. For example the one-way hypothesis of equality of all $\overline{F}_{i\cdot}$ (or that all A_i in (4.1.3) are zero) corresponds to $\mathbf{C} = (\mathbf{1}_{a-1}| - \mathbf{I}_{a-1})$, where $\mathbf{1}_a$ is an a-dimensional vector of 1's and $\mathbf{I}_a$ is the $a \times a$ identity matrix. For additional examples see Akritas, Arnold and Brunner (1997).

The rest of the chapter is organized as the following. We introduce the test statistic in Section 4.2. In Section 4.3, the asymptotic distributions of the test statistic under the null and local alternative hypotheses are presented. Numerical simulations are given in Section 4.4. Proofs of the main results are given in Section 4.5.1, while some auxiliary results and technical derivations are given in the Section 4.5.2

4.2 Construction of Test Statistics

Consider the ANCOVA model in the preceding section with one continuous covariate. Without loss of generality, we may assume that the covariate values are distinct, that in each group the observation pairs have been ordered by the covariate values and the current indices represent such ordering.

The basic idea for constructing a test statistic is to treat the covariate as a factor with many levels. Consider for simplicity a one-way ANCOVA design.

Then, the ANCOVA design is thought of as an $a \times N$ ANOVA design. Note that, with the suggested choices of G in (4.1.2), the main row effects in the hypothetical ANOVA design coincide with the main factor effects of the ANCOVA design if the X_{ij} are fixed (or for conditional analysis), and do so asymptotically if the X_{ij} are random. However, since there is only one observation per covariate value, this hypothetical two-way ANOVA design has at most one observation per cell and thus we cannot have a test statistic. To remedy this, we use smoothness assumptions and augment the cells with observations falling in a window around each covariate value. In particular, we first pool the covariates X_{ij} in ascending order and relabel them: $X_1, X_2, \ldots, X_N$. These form labels for the column levels in the hypothetical ANOVA design. For the i-th factor level (i.e. i-th row in the hypothetical ANOVA design), windows W_{ir} of size k with the center X_r are created, $r = 1, \ldots, N$. The window W_{ir} is formed from the k observations (Y_{ij}, X_{ij}) whose covariate values X_{ij} are nearest to X_r among $X_{i1}, \ldots, X_{in_i}$. By nearest to X_r we mean that their rank in $X_1, X_2, ..., X_N$ is closest to that of X_r. More explicitly, Y_{ij} belongs in the window W_{ir} if

$$|\widehat{G}_i(X_{ij}) - \widehat{G}_i(X_r)| \leq \frac{k-1}{2n_i}, \tag{4.2.1}$$

where $\widehat{G}_i(x) = (n_i)^{-1} \sum_{j=1}^{n_i} I\left(X_{ij} \leq x\right)$ is the empirical distribution function of the covariate in the i-th group. In what follows the windows W_{ir} will also be understood as sets containing the indices j of the covariate values in the i-th group that belong to that window, that is

$$W_{ir} = \left\{ j : \; I\left[|\widehat{G}_i(X_{ij}) - \widehat{G}_i(X_r)| \leq \frac{k-1}{2n_i} \right] \right\}. \tag{4.2.2}$$

We also note that $W_{i,X_{i'j'}}$ will be also used instead of W_{ir} if $X_{i'j'} = X_r$. For simplicity we may take k to be odd. To distinguish between the augmented ANOVA values and the truly observed data values, we label the t-th observation in the (i, r)-th cell of the augmented hypothetical two-way ANOVA Z_{irt}, more specifically, $Z_{irt} = Y_{ij}$ iff (4.2.1) is satisfied and $\sum_{l=1}^{n_i} I(X_{il} \leq X_{ij}, l \in W_{ir}) = t$.

This leads us to consider

$$\overline{\mathbf{Z}}_{..} = (\overline{Z}_{1..}, \overline{Z}_{2..}, \ldots, \overline{Z}_{a..})', \quad \text{where} \quad \overline{Z}_{i..} = \frac{1}{Nk} \sum_{r=1}^{N} \sum_{t=1}^{k} Z_{irt}, \tag{4.2.3}$$

as the random vector on which a statistic for testing the nonparametric hypothesis of no treatment effects can be based. Note that the dependence of $\overline{\mathbf{Z}}_{..}$ on k is not made explicit.

Remark 4.2.1. *The results of Bhattacharya (1974) indicate that even before augmentation, the observations in the same row are only conditionally independent (but not conditionally identically distributed).* Z_{irt} *are the so-called induced (or concomitant) order statistic.*

According to the discussion in Section 4.1, extension to multi-factor designs is immediately possible by thinking of i as enumerating the level combinations of several factors. For convenience, we will keep the one-way notation as representative for all designs. Recall that all possible hypotheses that we will consider are given by a contrast matrix $\mathbf{C}$. The actual test statistic for a hypothesis represented by contrast matrix $\mathbf{C}$ is of the form

$$Q_k(\mathbf{C}) = N(\mathbf{C}\overline{\mathbf{Z}}_{..})'(\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}')^{-1}(\mathbf{C}\overline{\mathbf{Z}}_{..}),$$

where k is the window size and $\widehat{\mathbf{V}}$ is a suitable estimator of the asymptotic covariance matrix of $\overline{\mathbf{Z}}_{..}$. In the next section it will be shown that $Q_k(\mathbf{C})$ has an asymptotic χ^2 distribution as $N \to \infty$, for both the fixed and random covariate cases. In all derivations the window size k remains fixed.

4.3 Main Results

For convenience, all the conditions required for the main results are stated here.

Assumption A1. For each i, $\widehat{\lambda}_i = \frac{n_i}{N} \to \lambda_i \in (0,1)$ as $n_i \to \infty$.

Assumption A2. The X_i are continuous random variables with cumulative distribution functions G_i and density function g_i, which have common bounded support S, $i = 1, \ldots, a$. The g_i are differentiable and are bounded away from 0 on S uniformly in i. In the fixed design case, the covariate are assumed to be a regular sequence generated from design density g_i in the sense of Sacks and Ylvisaker(1970), the design densities are assumed to have the same properties as above.

Assumption A3. The $E\left(Y_{ij}^4 | X_{ij} = x\right)$ are uniformly bounded in i, x.

Assumption A4. The $\sigma_i^2(x) = Var(Y_{ij}|X_{ij} = x)$, are bounded away from 0 on S uniformly in i, x.

Assumption A5. The $E(Y_{ij}|X_{ij} = x)$, are Lipschitz continuous in x.

The next two propositions are needed for the proof of the main results. The first proposition pertains to both random and fixed design.

Proposition 4.3.1. *Assume assumptions A.1-A.4, then, with $\overline{Z}_{i..}$ defined in (4.2.3),*

$$\left(N^{1/2}(\overline{Z}_{1..} - E(\overline{Z}_{1..}|\mathbf{X})), \ldots, N^{1/2}(\overline{Z}_{a..} - E(\overline{Z}_{a..}|\mathbf{X}))\right)' \to N_a(0, diag(c_1, \ldots, c_a)),$$

in distribution conditionally on $\mathbf{X}$, where defining $g = \sum_{i=1}^a \lambda_i g_i$,

$$\begin{aligned} c_i &= \frac{\lambda_i}{k^2} \int \sigma_i^2(x)\ dG_i(x) + \frac{k-1}{k^2} \int \sigma_i^2(x) \frac{g(x)}{g_i(x)}\ dG_i(x) \\ &\quad + \frac{k(k-1)}{\lambda_i k^2} \int \sigma_i^2(x) \frac{g^2(x)}{g_i^2(x)}\ dG_i(x). \end{aligned}$$

The proof of Proposition 4.3.1 is given in the next section.

Proposition 4.3.2. *Under assumptions A1, A2 for the random design case,*

$$N^{1/2}\left(\int y\ d\int F_{1x}(y)d(\widehat{G}(x)-G(x)),\ldots,\int y\ d\int F_{ax}(y)d(\widehat{G}(x)-G(x))\right) \to N_a(0,\mathbf{\Sigma}),$$

where $G(x) = E\left[\widehat{G}(x)\right]$, $\mathbf{\Sigma}$ *is an* $a \times a$ *matrix with diagonal elements* $\sigma^2_{1,i} = Var\left[E(Y_{ij}|X)\right]$ *and off-diagonal elements* $\sigma_{1,i_1i_2} = Cov\left(E(Y_{i_1j}|X), E(Y_{i_2j}|X)\right)$.

The proof of Proposition 4.3.2 follows by a straightforward application of the multivariate CLT.

Proposition 4.3.3. *Consistent estimators of* $\sigma^2_{1,i}$, σ_{1,i_1i_2} *and* c_i, *defined in Propositions 4.3.1, 4.3.2, are*

$$\widehat{\sigma}^2_{1,i} = \frac{1}{N}\sum_{r=1}^{N}\left(\overline{Z}_{ir.}-\overline{Z}_{i..}\right)^2,$$

$$\widehat{\sigma}_{1,i_1,i_2} = \frac{1}{N}\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\left(\overline{Z}_{i_1r.}-\overline{Z}_{i_1..}\right)\left(\overline{Z}_{i_2r.}-\overline{Z}_{i_2..}\right),$$

$$\widehat{c}_i = \frac{\widehat{\lambda}_i}{k^2n_i}\sum_{j=1}^{n_i}\widehat{\sigma}^2_i(X_{ij}) + \frac{k-1}{k^2n_i}\sum_{j=1}^{n_i}\widehat{\sigma}^2_i(X_{ij})\frac{n_i}{kN}\sum_{r=1}^{N} I\left[|\widehat{G}_i(X_{ij})-\widehat{G}_i(X_r)| \le \frac{k-1}{2n_i}\right]$$
$$+\frac{k(k-1)}{k^2}\frac{1}{\widehat{\lambda}_i n_i}\sum_{j=1}^{n_i}\widehat{\sigma}^2_i(X_{ij})\left\{\frac{n_i}{kN}\sum_{r=1}^{N} I\left[|\widehat{G}_i(X_{ij})-\widehat{G}_i(X_r)| \le \frac{k-1}{2n_i}\right]\right\}^2$$

where

$$\widehat{\sigma}^2_i(X_{ij}) = \frac{k}{k-1}\left\{\frac{1}{k}\sum_{l=1}^{n_i} Y^2_{il} I\left[|\widehat{G}_i(X_{il})-\widehat{G}_i(X_{ij})| \le \frac{k-1}{2n_i}\right]\right.$$
$$\left. -\left(\frac{1}{k}\sum_{l=1}^{n_i} Y_{il} I\left[|\widehat{G}_i(X_{il})-\widehat{G}_i(X_{ij})| \le \frac{k-1}{2n_i}\right]\right)^2\right\}.$$

The proof of Proposition 4.3.3 is not very difficult and thus it is omitted. We do remark, however, that the proof relies on the small bias (of order $1/n_i$) of the smoothed quantities. For example, $\overline{Z}_{ir.}$ is not a consistent estimator of $E(Y_{ij}|X_{ij} = X_r)$ due to the fact that the window size k does not tend to infinity. Additional insight into the estimator of c_i can be gained from

$$\frac{1}{k}\int I\left[|G_i(X_{ij}) - G_i(x)| \leq \frac{k-1}{2n_i}\right] dG(x) \quad = \quad \frac{g(X_{ij})}{n_i g_i(X_{ij})} + o_p(N^{-1/2})$$

which follows easily, first by a change of variable and then using Taylor expansion.

Theorem 4.3.4. *Let a be the total number of factor level combinations in an ANCOVA design with random covariate, and let assumptions A.1-A.5, for the random design case, hold. Then, under the null hypothesis (4.1.4),*

$$N(\mathbf{C}\overline{\mathbf{Z}}_{..})'(\mathbf{C}\hat{\mathbf{V}}\mathbf{C}')^{-1}(\mathbf{C}\overline{\mathbf{Z}}_{..}) \rightarrow \chi^2_\nu \quad \text{as N} \rightarrow \infty,$$

where $\mathbf{C}$ *is a full row-rank* $\nu \times a$ *contrast matrix, and* $\widehat{\mathbf{V}}$ *is an* $a \times a$ *matrix with diagonal elements* $\widehat{\sigma}^2_{1,i} + \hat{c}_i$ *and off-diagonal elements are* $\widehat{\sigma}_{1,i_1 i_2}$*, which are defined in Proposition 4.3.3.*

Theorem 4.3.5. *Let a be the total number of factor level combinations in an ANCOVA design with nonrandom covariate, and let assumptions A.1-A.5, for the fixed design case, hold. Then, under the null hypothesis (4.1.4),*

$$N(\mathbf{C}\overline{\mathbf{Z}}_{..})'(\mathbf{C}\widehat{\mathbf{V}}^*\mathbf{C}')^{-1}(\mathbf{C}\overline{\mathbf{Z}}_{..}) \rightarrow \chi^2_\nu \quad \text{as N} \rightarrow \infty,$$

where $\mathbf{C}$ *is a full row-rank* $\nu \times a$ *contrast matrix, and* $\widehat{\mathbf{V}}^* = diag(\hat{c}_1, \ldots, \hat{c}_a)$*, where the* $\widehat{c}_i$ *are defined in Proposition 4.3.3.*

Next, we consider the asymptotic behavior of the test statistic under local alternatives. A contiguous sequence of alternatives to the null hypothesis (4.1.4) is

$$F_{N,i}(y|x) = F_{N,ix}(y) = F_{ix}(y) + \frac{1}{\sqrt{N}} R_{ix}(y), \text{ for all } x,\ y, \text{ and } i = 1, \ldots, a, \quad (4.3.1)$$

where $F_{ix}(y)$, $i = 1, \ldots, a$, satisfy the null hypothesis in (4.1.4) and $R_{ix}(y)$, $i = 1, \ldots, a$, are functions of bounded variation measuring the departure from the null hypothesis.

The following theorem pertains to both the random and fixed design cases.

Theorem 4.3.6. *Assuming assumptions A.1-A.5, and consider the sequence of alternatives given in* (4.3.1). *It is further assumed that the $R_{ix}(y)$ are such that $\int y^2 dR_{ix}(y)$ are uniformly bounded for all i and x. Define $\delta_i = \int y d\left[\int R_{ix}(y)g(x)dx\right]$, $i = 1, \ldots, a$, and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_a)'$. Then, as $N \to \infty$, we have*

$$N(\mathbf{C}\overline{\mathbf{Z}}_{..})'(\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}')^{-1}(\mathbf{C}\overline{\mathbf{Z}}_{..}) \xrightarrow{d} \chi^2_{a-1}(\eta_1)\,, \quad \textit{if the covariates are random, and}$$
$$N(\mathbf{C}\overline{\mathbf{Z}}_{..})'(\mathbf{C}\widehat{\mathbf{V}}^*\mathbf{C}')^{-1}(\mathbf{C}\overline{\mathbf{Z}}_{..}) \xrightarrow{d} \chi^2_{a-1}(\eta_2)\,, \quad \textit{if the covariates are nonrandom}$$

where $\eta_1 = \boldsymbol{\delta}'\mathbf{C}'(\mathbf{C}\mathbf{V}\mathbf{C}')^{-1}\mathbf{C}\boldsymbol{\delta}$ and $\eta_2 = \boldsymbol{\delta}'\mathbf{C}'(\mathbf{C}\mathbf{V}^*\mathbf{C}')^{-1}\mathbf{C}\boldsymbol{\delta}$, with $\mathbf{V}$, $\mathbf{V}^\star$ being the covariance matrices that are estimated by $\widehat{\mathbf{V}}$, $\widehat{\mathbf{V}}^*$ given in Theorems 4.3.4, 4.3.5, respectively.

4.4 Numerical Simulations

The purpose of the simulation study in this section is to examine the performance of the proposed test statistic, in terms of the achieved type I error probability, under different window sizes, and to investigate its power properties. The R statistical package was used and all the results are based on 5000 runs.

In order to have k observations in each window, we adjust W_{ir} in the case the rank of X_r within group i is less than $k/2$ or greater than $n_i - k/2$. That is, if the rank of X_r within group i is less than $k/2$, W_{ir} is set to be $j = 1, \ldots, k$, and when greater than $n_i - k/2$, W_{ir} is set to be $j = (n_i - k), \ldots, n_i$.

A balanced design with three groups is used throughout, and the common group size will be denoted by n. In the first part of the simulation study we examine the level of the test for different values of k, for n =30, 50 and 80. In the second part we investigate the power properties using $k = 7$, which was one of the values that performed well in terms of type I error ($k = 5$ performed equally well), with 50 and 80 observations per group.

For the first part, the response variable and the covariate are generated under the following distributions:

Random Covariate

Case R1: In this case X_{ij} are iid $U(0, 1)$ for $i = 1, 2, 3$, $Y_{1j} = 2X_{1j} + \epsilon_{1j}$, $Y_{2j} = 2 - 2X_{2j} + \epsilon_{2j}$ and $Y_{3j} = 1 + \epsilon_{3j}$, where $\epsilon_{ij} \sim N(0, 1)$, independent.

Case R2: In this case X_{1j} are iid U(10,20), and with $B_2 \sim Beta(3, 3)$, $B_3 \sim Beta(5, 5)$, X_{2j} are iid distributed as $10 + 10B_2$, and X_{3j} are iid distributed as $10 + 10B_3$. Finally $Y_{ij} = 1 + .3X_{ij} + \epsilon_{ij}$, $i = 1, 2, 3$ and $j = 1, \ldots, n$, where $\epsilon_{ij} \sim N(0, 1)$, independent.

Fixed Covariate.

Case F1: The design points are generated by $x_{ij} = j/(n+1)$ for $i = 1, 2, 3$, $j = 1, \ldots, n$, while the responses are $Y_{1j} = x_{1j}/\overline{x} + \epsilon_{1j}$, $Y_{2j} = 2 - x_{2j}/\overline{x} + \epsilon_{2j}$ and $Y_{3j} = 1 + \epsilon_{3j}$ where $\overline{x}$ is the overall mean of the x_{ij}'s and $\epsilon_{ij} \sim N(0, 1)$, independent.

Case F2: Here the design points for each group are generated from $U(0,1)$ and are held fixed in all simulation runs. As a result, x_{ij}'s are different in each group and now, not equally spaced. Responses are generated as in case F1.

Case F3: Let b_{1j}, b_{2j} and b_{3j}, $j = 1, \ldots, n$, be the percentiles of $U(0,1)$, $Beta(3,3)$ and $Beta(5,5)$, respectively, such that they have equal probability spacings, and set $x_{ij} = 10 + 10b_{ij}$, $i = 1,2,3$, $j = 1, \ldots, n$. The response is generated as $Y_{ij} = 1 + .3X_{ij} + \epsilon_{ij}$, $i = 1,2,3$ and $j = 1, \ldots, n$, where $\epsilon_{ij} \sim N(0,1)$, independent.

The achieved alpha levels for the two random covariate cases in Table 4.1 and for the three fixed covariate cases are summarized in Table 4.2. In all cases the nominal level is $\alpha = 0.05$, the results for $\alpha = 0.01$ are similar and are omitted.

	Case R1			Case R2		
		n			n	
k	30	50	80	30	50	80
5	0.0474	0.0452	0.0412	0.0444	0.0452	0.0376
7	0.0526	0.0440	0.0448	0.0492	0.0424	0.0438
9	0.0544	0.0546	0.0406	0.0470	0.0450	0.0434
11	0.0602	0.0476	0.0456	0.0422	0.0490	0.0488
13	0.0630	0.0530	0.0482	0.0484	0.0438	0.0538
15	0.0616	0.0516	0.0510	0.0482	0.0458	0.0440
17	0.0568	0.0536	0.0462	0.0476	0.0492	0.0442

Table 4.1: Choice of k for cases with random X and no group effect.

In the case of a random covariate, we can see that for Case R1, the test is getting more liberal for smaller sample sizes, and for large values of k. For

	Case F1.1			Case F1.2			Case F2		
	n			n			n		
k	30	50	80	30	50	80	30	50	80
5	0.0460	0.0476	0.0458	0.0486	0.0450	0.0460	0.0480	0.0472	0.0476
7	0.0460	0.0470	0.0478	0.0462	0.0476	0.0508	0.0448	0.0486	0.0480
9	0.0482	0.0496	0.0482	0.0456	0.0492	0.0504	0.0454	0.0468	0.0480
11	0.0470	0.0496	0.0488	0.0470	0.0490	0.0492	0.0416	0.0478	0.0464
13	0.0476	0.0498	0.0496	0.0460	0.0486	0.0506	0.0390	0.0454	0.0484
15	0.0460	0.0506	0.0504	0.0462	0.0482	0.0514	0.0368	0.0442	0.0468
17	0.0436	0.0470	0.0518	0.0448	0.0468	0.0478	0.0336	0.0426	0.0466

Table 4.2: Choice of k for cases with fixed X and no group effect.

$k = 7$ the results obtained for each sample size are satisfactory. In Case R2, we can see that the simulated alpha values are pretty robust with respect to sample size and k.

From Table 4.2, we can see that for the first two cases, F1.1 and F1.2, the results are similar and that the achieved alpha level is pretty robust with respect to the choice of k. Furthermore, we observe that for smaller samples, $n = 30$ or 50, smaller values of k, $k = 7$ or 9 work slightly better than larger values. For sample size 80, $k = 13$ and 15 seem to be the optimal choices, however the results for smaller values of k are very satisfactory. In Case F2, we observe similar trends.

Overall, the simulated alpha levels tend to be conservative; however, they are pretty satisfactory.

Finally, we perform simulations to investigate the power properties of the proposed test. In this case, the response is simulated independently of the covariate. In the case of a random covariate, X is simulated from $U(0, 1)$; while in the case of a fixed covariate, the design points for each group are equally spaced on (0,1). The response is generated from a standard normal or a lognormal distribution. For comparison purposes, we also report the results from the classical analysis of covariance, denoted by CF on the table, using equal slopes in each group.

We are using linear factor effects, and as expected, the classical analysis of variance is performing well. Recall from the previous results we have seen that $k = 7$ lead to satisfactory alpha levels regardless the sample size. Therefore, we used $k = 7$ for both sample sizes that are used in this study ($n = 50$ and $n = 80$). From Tables 4.3 and 4.4, we can see that our method, denoted by New-χ^2 on the table, behaves similar to the ANCOVA method for both covariate cases. Our method appears to have a slightly lower power for the normal case and slightly higher power in the lognormal case.

Random: $X \sim U(0,1)$					
		$\epsilon_{ij} \sim N(0,1)$		$\epsilon_{ij} \sim exp(N(0,1))$	
	τ	New-χ^2	CF	New-χ^2	CF
	0.0	0.0464	0.0460	0.0516	0.0372
$n = 50$	0.1	0.1228	0.1230	0.0772	0.0648
$k = 7$	0.2	0.4022	0.4086	0.1578	0.1338
	0.3	0.7434	0.7574	0.2878	0.2598
	0.4	0.9512	0.9560	0.4608	0.4398
	0.5	0.9966	0.9964	0.6516	0.6240
	0.0	0.0442	0.0454	0.0436	0.0418
$n = 80$	0.1	0.1748	0.1880	0.0846	0.0736
$k = 7$	0.2	0.5942	0.6104	0.2026	0.1952
	0.3	0.9226	0.9294	0.4000	0.3890
	0.4	0.9942	0.9960	0.6140	0.5908
	0.5	1.0000	1.0000	0.7954	0.7780

Table 4.3: Power Study; 3 groups of equal size $n = 50$ and $n = 80$ and $\alpha = 0.05$. $Y_{ij} = (i-1)\tau + \epsilon_{ij}$

Fixed: $x_{ij} = j/(n+1)$					
		$\epsilon_{ij} \sim N(0,1)$		$\epsilon_{ij} \sim exp(N(0,1))$	
	τ	New-χ^2	CF	New-χ^2	CF
	0.0	0.0468	0.0524	0.0472	0.0396
$n = 50$	0.1	0.1140	0.1316	0.0668	0.0618
$k = 7$	0.2	0.3640	0.4072	0.1386	0.1328
	0.3	0.6858	0.7466	0.2608	0.2612
	0.4	0.9228	0.9486	0.4320	0.4446
	0.5	0.9914	0.9950	0.5970	0.6022
	0.0	0.0450	0.0486	0.0418	0.0422
$n = 80$	0.1	0.1592	0.1916	0.0768	0.0754
$k = 7$	0.2	0.5438	0.6138	0.1838	0.1968
	0.3	0.8876	0.9320	0.3662	0.3804
	0.4	0.9928	0.9968	0.5780	0.5892
	0.5	1.0000	1.0000	0.7630	0.7742

Table 4.4: Power Study; 3 groups of equal size $n = 50$ and $n = 80$ and $\alpha = 0.05$. $Y_{ij} = (i-1)\tau + \epsilon_{ij}$

4.5 Proofs

In the first part of this section presents the proofs of the main results given in Section 4.3 and the second part presents some useful lemmas and the proof of a step in the proof of Proposition 4.3.1 containing auxiliary results.

4.5.1 Main Proofs

Proof of Proposition 4.3.1. By the independence of the $\overline{Z}_{i\cdot\cdot}$ it is sufficient to show convergence of each coordinate $\widehat{T}_i = \overline{Z}_{i\cdot\cdot} - E(\overline{Z}_{i\cdot\cdot}|\mathbf{X})$ to the corresponding marginal distribution. Thus, using $\overline{Z}_{i\cdot\cdot} = (Nk)^{-1}\sum_{r=1}^{N}\sum_{j=1}^{n_i} Y_{ij} I(j \in W_{ir})$, we need to show that

$$N^{1/2}\widehat{T}_i = \frac{N^{1/2}}{Nk}\sum_{r=1}^{N}\sum_{j=1}^{n_i}(Y_{ij} - E(Y_{ij}|\mathbf{X}))I(j \in W_{ir}) = N^{1/2}\sum_{j=1}^{n_i} t_{ij} \to N(0, c_i), \quad (4.5.1)$$

in distribution conditionally on $\mathbf{X}$, where

$$t_{ij} = (Y_{ij} - E(Y_{ij}|\mathbf{X})) \left(\frac{1}{Nk} \sum_{r=1}^{N} I(j \in W_{ir}) \right).$$

Relation (4.5.1) will be shown if we show that

$$Var\left(N^{1/2}\widehat{T}_i|\mathbf{X}\right) \xrightarrow{p} c_i, \quad \text{and} \quad \frac{\widehat{T}_i}{\left[Var\left(\widehat{T}_i|\mathbf{X}\right)\right]^{1/2}} \to N(0,1), \qquad (4.5.2)$$

in distribution, conditionally on $\mathbf{X}$. For both parts we need to find a suitable expression for $Var\left(\widehat{T}_i|\mathbf{X}\right)$. Noting that conditionally on $\mathbf{X}$, t_{ij} are independent, we have

$$\begin{aligned} Var\left(\widehat{T}_i|\mathbf{X}\right) &= \sum_{j=1}^{n_i} Var(t_{ij}|\mathbf{X}) \\ &= \sum_{j=1}^{n_i} E((Y_{ij} - E(Y_{ij}|\mathbf{X}))^2|\mathbf{X}) \left(\frac{1}{Nk} \sum_{r=1}^{N} I(j \in W_{ir}) \right)^2. \end{aligned}$$

Next, write $\sum_{r=1}^{N} I(j \in W_{ir}) = \sum_{i'=1}^{a} \sum_{j'=1}^{n_i} I(j \in W_{i,X_{i'j'}})$, the number of windows that contain each j. Since the indexes in W_{ir} correspond only to covariate values from group i, it follows that $\sum_{j'=1}^{n_i} I(j \in W_{i,X_{ij'}}) = k$. However, due to the randomness of the design points the count cannot be exact when the window is centered at a covariate value that does not belong in group i. Thus we will employ Lemmas 4.5.1, 4.5.2, which provide sufficient tools for proving (4.5.2). In particular, Lemma 4.5.2 implies that, for all $0 < \delta < 1$,

$$k \leq \sum_{r=1}^{N} I(j \in W_{ir}) = k + \sum_{i' \neq i}^{a} \sum_{j'=1}^{n_i'} I(j \in W_{i,X_{i'j'}}) \leq CN^{\delta},$$

for some positive constant C. For the second part of (4.5.2), we use the above lower and upper bounds to obtain

$$\frac{\max_{1 \leq j \leq n_i} Var(t_{ij}|\mathbf{X})}{\sum_{j=1}^{n_i} Var(t_{ij}|\mathbf{X})} \leq C^* N^{2\delta - 1} \to 0,$$

for some positive constant $C^* > 0$, $\forall\ 0 < \delta < 1/2$. Thus, the condition of Theorem 2.7.4 (Lehmann, 1998) holds and the second part of (4.5.2) is shown.

The first part of (4.5.2) will follow by showing

$$(i)\ E(\xi_i) \to c_i, \quad \text{and} \quad (ii)\ \xi_i - E(\xi_i) \xrightarrow{p} 0, \tag{4.5.3}$$

where c_i is defined in Proposition 4.3.1, and

$$\xi_i = Var(\sqrt{N}\widehat{T}_i|\mathbf{X}) = \frac{1}{Nk^2}\sum_{j=1}^{n_i}\sigma_i^2(X_{ij})\left[\sum_{r=1}^{N} I(j \in W_{ir})\right]^2. \tag{4.5.4}$$

Consider first (4.5.3)(i). Expanding the square in the expression for ξ_i we have

$$\begin{aligned} E(\xi_i) &= \frac{n_i}{Nk^2}E\left[\sigma_i^2(X_{ij})\sum_{r=1}^{N} I(j \in W_{ir})\right] \\ &\quad + \frac{n_i}{Nk^2}E\left[\sigma_i^2(X_{ij})\sum_{r_1 \neq r_2}^{N} I(j \in W_{ir_1} \cap W_{ir_2})\right] \\ &= \frac{n_i}{Nk^2}[Q_1 + Q_2], \quad \text{say,} \end{aligned} \tag{4.5.5}$$

where the dependence of Q_1 and Q_2 on i is suppressed. Write

$$\begin{aligned} Q_1 &= E\left\{E\left[\sigma_i^2(X_{ij})\sum_{r=1}^{N} I(j \in W_{ir})|\mathbf{X}_i\right]\right\} \\ &= E\left\{\sigma_i^2(X_{ij})E\left[\sum_{\substack{X_r \in \\ group\ i}} I(j \in W_{ir})|\mathbf{X}_i\right] + \sigma_i^2(X_{ij})E\left[\sum_{\substack{X_r \notin \\ group\ i}} I(j \in W_{ir})|\mathbf{X}_i\right]\right\} \\ &= E\left\{\sigma_i^2(X_{ij})Q_{11,j}(\mathbf{X}_i) + \sigma_i^2(X_{ij})Q_{12,j}(\mathbf{X}_i)\right\}, \end{aligned} \tag{4.5.6}$$

where the definition of $Q_{11,j}(\mathbf{X}_i)$ and $Q_{12,j}(\mathbf{X}_i)$ should be clear from the context. It is obvious that $Q_{11,j}(\mathbf{X}_i) = k$. Using (4.5.26), and two Taylor expansions we have

$$Q_{12,j}(\mathbf{X}_i) = \sum_{i_1 \neq i} n_{i_1} \int_{X_{i,j-p}}^{X_{i,j+p}} g_{i_1}(x)dx = \sum_{i_1 \neq i} n_{i_1} \int_{X_{ij}-L_{ij}}^{X_{ij}+U_{ij}} g_{i_1}(x)dx$$

$$= \sum_{i_1 \neq i} n_{i_1} \left[g_{i_1}(X_{ij})(U_{ij} + L_{ij}) + O(U_{ij}^2) + O(L_{ij}^2) \right]$$

$$= \sum_{i_1 \neq i} n_{i_1} \left[g_{i_1}(X_{ij}) \frac{G_i(X_{ij} + U_{ij}) - G_i(X_{ij} - L_{ij})}{g_i(X_{ij})} + O(U_{ij}^2) + O(L_{ij}^2) \right], \quad (4.5.7)$$

where $p = (k-1)/2$ and U_{ij}, L_{ij} are the upper, lower p spacings from X_{ij}. Moreover, if $D_{ij}(k-1) = G_i(X_{ij} + U_{ij}) - G_i(X_{ij} - L_{ij})$, then results of Pyke (1965) imply

$$E\left[D_{ij}(k-1)|X_{ij}\right] = \frac{k-1}{n_i+1}, \quad E\left[D_{ij}(k-1)^2|X_{ij}\right] = \frac{k(k-1)}{(n_i+1)(n_1+2)}. \quad (4.5.8)$$

Note that the second relation in (4.5.8) implies that $E(U_{ij}^2|X_{ij}) = O_p(N^{-2}) = E(L_{ij}^2|X_{ij})$. Combining the above with (4.5.6) we get

$$Q_1 = kE[\sigma_i^2(X_{ij})] + E\{\sigma_i^2(X_{ij})E[Q_{12,j}(\mathbf{X}_i)|X_{ij}]\}$$

$$= kE[\sigma_i^2(X_{ij})] + E\left\{\sigma_i^2(X_{ij}) \sum_{i_1 \neq i} \frac{n_{i_1} g_{i_1}(X_{ij})}{g_i(X_{ij})} \frac{k-1}{n_i+1} + O_p(N^{-1})\right\}. \quad (4.5.9)$$

Now consider Q_2. Notice that

$$E\left[\sum_{r_1 \neq r_2}^{N} I(j \in W_{ir_1} \cap W_{ir_2})|\mathbf{X}_i\right] = Q_{21,j}(\mathbf{X}_i) + Q_{22,j}(\mathbf{X}_i) + Q_{23,j}(\mathbf{X}_i),$$

where $Q_{21,j}(\mathbf{X}_i)$, $Q_{22,j}(\mathbf{X}_i)$ and $Q_{23,j}(\mathbf{X}_i)$ correspond to the sums when $X_{r_1}, X_{r_2} \in$ *group* i, $X_{r_1} \in$ *group* $i, X_{r_2} \notin$ *group* i or vice versa, and $X_{r_1}, X_{r_2} \notin$ *group* i, respectively. It is obvious that $Q_{21}(X_{ij}) = k(k-1)$. Next,

$$Q_{22,j}(\mathbf{X}_i) = 2E\left[\sum_{X_{r_1} \in group\ i} I(X_{ij} \in W_{ir_1}) \sum_{X_{r_2} \notin group\ i} I(X_{ij} \in W_{ir_2})|\mathbf{X}_i\right]$$

$$= 2kQ_{12,j}(\mathbf{X}_i).$$

Finally, using (4.5.26) it is easily seen that when $(i_1, j_1) \neq (i_2, j_2)$, $I(j \in W_{i,X_{i_1 j_1}})$, $I(j \in W_{i,X_{i_2 j_2}})$ are conditionally independent given $\mathbf{X}_i$. Thus, since $I(j \in W_{ir_1} \cap W_{ir_2}) = I(j \in W_{i,X_{i_1 j_1}}) I(j \in W_{i,X_{i_2 j_2}})$, it follows

$$\begin{aligned} Q_{23,j}(\mathbf{X}_i) &= \sum_{\substack{r_1 \neq r_2; X_{r_1} \notin group\ i, \\ X_{r_2} \notin group\ i}} E\left[I(j \in W_{ir_1})|\mathbf{X}_i\right] E\left[I(j \in W_{ir_2})|\mathbf{X}_i\right] \\ &= Q_{12,j}(\mathbf{X}_i)^2 - \sum_{X_r \notin group\ i} E[I(j \in W_{i,X_r})|\mathbf{X}_i]^2. \qquad (4.5.10) \end{aligned}$$

Using two Taylor expansions as was done in (4.5.7) and the results of Pyke (1965) stated in (4.5.8), we have that $E\left\{\sum_{X_r \notin group\ i} E[I(j \in W_{i,X_r})|\mathbf{X}_i]^2\right\} = O(N^{-1})$. Combining the above with the definition of Q_2, we have

$$Q_2 = E\left\{\sigma_i^2(X_{ij})[k(k-1) + 2kQ_{12,j}(\mathbf{X}_i) + Q_{12,j}(\mathbf{X}_i)^2]\right\} + O(N^{-1}). \quad (4.5.11)$$

Since $E\left[\sigma_i^2(X_{ij})Q_{12,j}(\mathbf{X}_i)\right]$ was evaluated above, see (4.5.9), it remains to evaluate $E\left\{\sigma_i^2(X_{ij})\ \ Q_{12,j}(\mathbf{X}_i)^2\right\} = E\left\{\sigma_i^2(X_{ij})E\left[Q_{12,j}(\mathbf{X}_i)^2|X_{ij}\right]\right\}$. Using (4.5.7) and (4.5.8) we have

$$\begin{aligned} & E\left\{Q_{12,j}(\mathbf{X}_i)^2|X_{ij}\right\} \\ = \ & E\left\{\sum_{i_1,i_2 \neq i} n_{i_1} n_{i_2} \left[g_{i_1}(X_{ij}) \frac{G_i(X_{ij}+U_{ij}) - G_i(X_{ij}-L_{ij})}{g_i(X_{ij})} + O(U_{ij}^2) + O(L_{ij}^2)\right]\right. \\ & \left. \times \left[g_{i_2}(X_{ij}) \frac{G_i(X_{ij}+U_{ij}) - G_i(X_{ij}-L_{ij})}{g_i(X_{ij})} + O(U_{ij}^2) + O(L_{ij}^2)\right] |X_{ij}\right\} \\ = \ & \sum_{i_1,i_2 \neq i} \left[\frac{n_{i_1} n_{i_2} g_{i_1}(X_{ij}) g_{i_2}(X_{ij})}{g_i^2(X_{ij})} \frac{k(k-1)}{(n_i+1)(n_i+2)} + O_p(N^{-1})\right]. \qquad (4.5.12) \end{aligned}$$

Using (4.5.5) and the subsequent derivations we have

$$\begin{aligned} E(\xi_i) &= E\left[\lambda_i\sigma_i^2(X_{ij}) + \frac{(2k+1)(k-1)}{k^2}\sigma_i^2(X_{ij})\sum_{i_1\neq i}\frac{\lambda_{i_1}g_{i_1}(X_{ij})}{g_i(X_{ij})}\right. \\ &\quad \left. +\frac{k(k-1)}{k^2}\;\sigma_i^2(X_{ij})\sum_{i_1\neq i}\sum_{i_2\neq i}\frac{\lambda_{i_1}\lambda_{i_2}g_{i_1}(X_{ij})g_{i_2}(X_{ij})}{\lambda_i g_i^2(X_{ij})}\right] + o(1) \\ &= E\left[\lambda_i\sigma_i^2(X_{ij}) + \frac{(2k+1)(k-1)}{k^2}\sigma_i^2(X_{ij})\frac{g(X_{ij})-\lambda_i g_i(X_{ij})}{g_i(X_{ij})}\right. \\ &\quad \left. +\frac{k(k-1)}{k^2}\;\sigma_i^2(X_{ij})\frac{[g(X_{ij})-\lambda_i g_i(X_{ij})]^2}{\lambda_i g_i^2(X_{ij})}\right] + o(1) \\ &= \frac{\lambda_i}{k^2}\int\sigma_i^2(x)g_i(x)dx + \frac{k-1}{k^2}\int\sigma_i^2(x)g(x)dx \\ &\quad +\frac{k(k-1)}{\lambda_i k^2}\int\sigma_i^2(x)\frac{g^2(x)}{g_i(x)}dx + o(1) \\ &= c_i + o(1). \end{aligned}$$

Finally, we need to prove (4.5.3)(ii). This is deferred to Section 4.5.2. Thus the first part of (4.5.2) is also shown and this completes the proof.□

Proof of Theorem 4.3.4. To obtain the asymptotic χ^2 limiting result, we will first establish the asymptotic normality, under $H_0(A)$, of $C(\overline{Z}_{1..},\ldots,\overline{Z}_{a..})'$. Because under $H_0(A)$,

$$\mathbf{C}\left(\int yd\left[\int F_{1x}(y)dG(x)\right],\ldots,\int yd\left[\int F_{ax}(y)dG(x)\right]\right)' = \mathbf{0}_{a\times 1}, \quad (4.5.13)$$

it suffices consider the asymptotic normality of

$$\mathbf{C}\left(\overline{Z}_{1..} - \int yd\left[\int F_{1x}(y)dG(x)\right],\ldots,\overline{Z}_{a..} - \int yd\left[\int F_{ax}(y)dG(x)\right]\right)',$$

or simply of $\left(\overline{Z}_{1..} - \int yd\left[\int F_{1x}(y)dG(x)\right],\ldots,\overline{Z}_{a..} - \int yd\left[\int F_{ax}(y)dG(x)\right]\right)'$.

The i-th coordinate of this random vector decomposes in two terms

$$\begin{aligned} & \overline{Z}_{i..} - \int yd\left[\int F_{ix}(y)dG(x)\right] \\ = \; & \overline{Z}_{i..} - E(\overline{Z}_{i..}|\mathbf{X}) + E(\overline{Z}_{i..}|\mathbf{X}) - \int yd\left[\int F_{ix}(y)dG(x)\right] \\ = \; & \overline{Z}_{i..} - E(\overline{Z}_{i..}|\mathbf{X}) + \int yd\left[\int F_{ix}(y)d(\widehat{G}(x) - G(x))\right] + o_p(N^{-1/2}), \end{aligned} \tag{4.5.14}$$

where the last equality follows from

$$\begin{aligned} E(\overline{Z}_{i..}|\mathbf{X}) &= \frac{1}{N}\sum_{r=1}^{N} E\left[\frac{1}{k}\sum_{t=1}^{k} Z_{irt}|\mathbf{X}\right] = \frac{1}{N}\sum_{r=1}^{N} E\left[\frac{1}{k}\sum_{j=1}^{n_i} Y_{ij} I(j \in W_{ir})|\mathbf{X}\right] \\ &= \frac{1}{N}\sum_{r=1}^{N}\left[\frac{1}{k}\sum_{j=1}^{n_i} E(Y_{ij}|X_{ij}) I(j \in W_{ir})\right] = \frac{1}{N}\sum_{r=1}^{N} E(Y_{ij}|X_r) + O_p(\frac{1}{N}) \\ &= \int E(Y_{ij}|x)d\widehat{G}(x) + O_p(\frac{1}{N}) = \int yd\left[\int F_{ix}(y)d\widehat{G}(x)\right] + O_p(\frac{1}{N}), \end{aligned} \tag{4.5.15}$$

Notice that the fourth equality in (4.5.15) follows by an application of Lemma A.3 of Akritas and Wang (2002), using Assumption A5. According to decomposition (4.5.14) we have $\sqrt{N}\left(\overline{Z}_{1..} - \int yd\int F_{1x}(y)dG(x), ..., \overline{Z}_{a..} - \int yd\int F_{ax}(y)dG(x)\right)'$ $= H_1(\mathbf{X}, \mathbf{Y}) + H_2(\mathbf{X})$, where

$$H_1(\mathbf{X}, \mathbf{Y}) = \sqrt{N}\left(\overline{Z}_{1..} - E(\overline{Z}_{1..}|\mathbf{X}), \dots, \overline{Z}_{a..} - E(\overline{Z}_{a..}|\mathbf{X})\right)',$$

$$H_2(\mathbf{X}) = \sqrt{N}\left(\int yd\int F_{1x}(y)d(\widehat{G}(x) - G(x)), ..., \int yd\int F_{ax}(y)d(\widehat{G}(x) - G(x))\right)'.$$

Using Lemma 4.5.3, Proposition 4.3.1 and Proposition 4.3.2 we have that under $H_0(A)$

$$N^{1/2}\left(\overline{Z}_{1..} - \int yd\left[\int F_{1x}(y)dG(x)\right], \dots, \overline{Z}_{a..} - \int yd\left[\int F_{ax}(y)dG(x)\right]\right)' \to N_a(\mathbf{0}, \mathbf{V}),$$

as $N \to \infty$, where $\mathbf{V} = \text{diag}(c_1, \ldots, c_a) + \boldsymbol{\Sigma}$. Finally, we obtain the asymptotic χ^2 distribution in Theorem 4.3.4 using the fact that $\mathbf{C}$ is a full row-rank $\nu \times a$ matrix, and that $\widehat{\mathbf{V}}$ is a consistent estimator of $\mathbf{V}$, see Proposition 4.3.3. □

Proof of Theorem 4.3.5. Notice that in the fixed covariate case, under $H_0(A)$ we have,

$$\mathbf{C}\left(\int yd\left[\int F_{i_1x}(y)d\widehat{G}(x)\right], \ldots, \int yd\left[\int F_{i_ax}(y)d\widehat{G}(x)\right]\right)' = \mathbf{0}_{a\times 1}, \quad (4.5.16)$$

Furthermore, form (4.5.15) we have

$$\overline{Z}_{i..} - \int yd\left[\int F_{ix}(y)d\widehat{G}(x)\right] = \left[\overline{Z}_{i..} - E(\overline{Z}_{i..}|\mathbf{X})\right] + O_p(N^{-1}).$$

Thus, applying Proposition 4.3.1 we have

$$N^{1/2}\left(\overline{Z}_{1..} - \int yd\left[\int F_{i_1x}(y)d\widehat{G}(x)\right], \ldots, \overline{Z}_{a..} - \int yd\left[\int F_{i_ax}(y)d\widehat{G}(x)\right]\right)' \to N_a(\mathbf{0}, \mathbf{V}^*),$$

as $N \to \infty$, where $\mathbf{V}^* = \text{diag}(c_1, \ldots, c_a)$.

The result of Theorem 4.3.5 follows using (4.5.16), and that $\widehat{\mathbf{V}}^*$ is a consistent estimator of $\mathbf{V}^*$, see Proposition 4.3.3. □

Proof of Theorem 4.3.6. Write $\mathbf{C}\overline{\mathbf{Z}}_{..}$ as

$$\begin{aligned}
&\mathbf{C}\left(\overline{Z}_{1..} - \int yd\left[\int F_{N,i_1x}(y)dG(x)\right], \ldots, \overline{Z}_{a..} - \int yd\left[\int F_{N,i_ax}(y)dG(x)\right]\right)' \\
&\quad + \mathbf{C}\left(\int yd\left[\int F_{N,i_1x}(y)dG(x)\right], \ldots, \int yd\left[\int F_{N,i_ax}(y)dG(x)\right]\right)' \\
&= \mathbf{C}\left(\overline{Z}_{1..} - \int yd\left[\int F_{N,i_1x}(y)dG(x)\right], \ldots, \overline{Z}_{a..} - \int yd\left[\int F_{N,i_ax}(y)dG(x)\right]\right)' \\
&\quad + N^{-1/2}\mathbf{C}\boldsymbol{\delta}' + o_p(N^{-1/2}). \quad (4.5.17)
\end{aligned}$$

Thus, in order to derive the results in Theorem 4.3.6, we will first consider the asymptotic distribution of

$$N^{1/2}\left(\overline{Z}_{1..} - \int yd\left[\int F_{N,i_1x}(y)dG(x)\right], \ldots, \overline{Z}_{a..} - \int yd\left[\int F_{N,i_ax}(y)dG(x)\right]\right)'.$$

Let $E_N(\cdot)$ denote the expectation under $F_{N,ix}$ defined in (4.3.1). Then, following similar steps as in (4.5.15),

$$E_N(\overline{Z}_{i..}|\mathbf{X}) = \int yd\left[\int F_{N,ix}(y)d\widehat{G}(x)\right] + o_p(N^{-1/2}). \qquad (4.5.18)$$

Hence, by adding and subtracting $E_N(\overline{Z}_{i..}|\mathbf{X})$ we can write

$$\begin{aligned} & \overline{Z}_{i..} - \int yd\left[\int F_{N,ix}(y)dG(x)\right] \\ = \; & \overline{Z}_{i..} - E_N(\overline{Z}_{i..}|\mathbf{X}) + \int yd\int F_{N,ix}(y)d(\widehat{G}(x) - G(x)) + o_p(N^{-1/2}). \end{aligned} \qquad (4.5.19)$$

We will first show that

$$N^{1/2}\left(\overline{Z}_{1..} - E_N(\overline{Z}_{1..}|\mathbf{X}), ..., \overline{Z}_{a..} - E_N(\overline{Z}_{a..}|\mathbf{X})\right) \rightarrow N_a(\mathbf{0}, diag(c_1, ..., c_a)), \quad (4.5.20)$$

as $N \rightarrow \infty$, conditionally on $\mathbf{X}$ where c_i are defined in Proposition 4.3.1. Note that by the independence of the $\overline{Z}_{i..}$, to prove (4.5.20) it is sufficient to show convergence of each coordinate $\widehat{T}_{N,i} = \overline{Z}_{i..} - E_N(\overline{Z}_{i..}|\mathbf{X})$ to the corresponding marginal distribution. Let $Var_N(.)$ to denote the variance under $F_{N,ix}$ defined in (4.3.1). Following the same steps as in the proof of the second part of (4.5.2), we can show that the asymptotic distribution of $\widehat{T}_{N,i}/Var_N\left(\widehat{T}_{N,i}|\mathbf{X}\right)$ conditioned on $\mathbf{X}$ is $N(0,1)$. Hence, to prove (4.5.20), it suffices to show that $\xi_{N,i} = Var_N\left(N^{1/2}\widehat{T}_{N,i}|\mathbf{X}\right) \xrightarrow{p} c_i$. To do so, we will show that

$$(i)\ E(\xi_{N,i}) \rightarrow c_i \quad \text{and} \quad (ii)\ \xi_{N,i} - E(\xi_{N,i}) \xrightarrow{p} 0. \qquad (4.5.21)$$

Writing $\widehat{T}_{N,i} = \sum_{r=1}^{N}\sum_{j=1}^{n_1}\left[Y_{ij} - E_N(X_{ij})\right] I(j \in W_{ir})$, using (4.5.23) we have that

$$
\begin{aligned}
E(\xi_{N,i}) &= \frac{1}{Nk^2}\sum_{j=1}^{n_i} E\left\{\sigma_{N,i}^2(X_{ij})\left[\sum_{r=1}^{N} I(j\in W_{ir})\right]^2\right\} \\
&= \frac{n_i}{Nk^2}E\left[\sigma_{N,i}^2(X_{ij})E\left\{\left(\sum_{r=1}^{N} I(j\in W_{ir})\right)^2 |\mathbf{X}_i\right\}\right] \\
&= \frac{n_i}{Nk^2}E\left[\sigma_{i}^2(X_{ij})E\left\{\left(\sum_{r=1}^{N} I(j\in W_{ir})\right)^2 |\mathbf{X}_i\right\}\right] + O(N^{-1/2}), \\
&= E(\xi_i) + O(N^{-1/2}) \qquad (4.5.22)
\end{aligned}
$$

where ξ_i is defined in (4.5.4). For the third equality we used

$$
\begin{aligned}
&\sigma_{N,i}^2(X_{ij}) = Var_N(Y_{ij}|X_{ij}) = E_N(Y_{ij}^2|X_{ij}) - [E_N(Y_{ij}|X_{ij})]^2 \\
&= \int y^2 dF_{ix}(y) + \frac{1}{\sqrt{N}}\int y^2 dR_{ix}(y) - \left[\int y dF_{ix}(y)\right]^2 - \left[\frac{1}{\sqrt{N}}\int y dR_{ix}(y)\right]^2 \\
&\quad -\frac{2}{\sqrt{N}}\int y dF_{ix}(y)\int y dR_{ix}(y) \\
&= \sigma_i^2(X_{ij}) + O_p(N^{-1/2}), \qquad (4.5.23)
\end{aligned}
$$

where $\sigma_{N,i}^2(X_{ij})$ is the conditional variance under the sequence of alternatives defined in (4.3.1), and $\sigma_i^2(X_{ij})$ under the null, and the fact that $E\left\{\left(\sum_{r=1}^{N} I(j\in W_{ir})\right)^2\right\}$ is bounded, (see proof of Proposition 4.3.1). Finally, a combination of (4.5.22) and (4.5.3)(i) completes the proof of (4.5.21)(i). The proof of (4.5.21)(ii) follows by steps similar to those of the proof of (4.5.3)(ii). This completes the proof of (4.5.20).

Next, we will show that

$$
N^{1/2}\left(\int yd\int F_{N,1x}(y)d(\widehat{G}(x)-G(x)),\ldots,\int yd\int F_{N,ax}(y)d(\widehat{G}(x)-G(x))\right)' \to N_a(\mathbf{0},\mathbf{\Sigma}), \quad (4.5.24)
$$

as $N \to \infty$, where $G(x)$ and $\mathbf{\Sigma}$ are defined in Proposition 4.3.2. Noting that

$$\begin{aligned}
&\int yd \int F_{N,ix}(y)d(\widehat{G}(x) - G(x)) \\
&= \int yd \int F_{ix}(y)d(\widehat{G}(x) - G(x)) + \frac{1}{\sqrt{N}} \int \left[\int ydR_{ix}(y) \right] d(\widehat{G}(x) - G(x)) \\
&= yd \int F_{ix}(y)d(\widehat{G}(x) - G(x)) + o_p(1),
\end{aligned}$$

since $\int ydR_{ix}(y)$ is bounded and from Assumption A2 we have that $G(x)$ has bounded support. Thus, (4.5.24) follows using the result in Proposition 4.3.2.

Now using (4.5.20), (4.5.24) and Lemma 4.5.3, we have that the right hand side of (4.5.19) is asymptotically normally distributed with mean zero and variance $\mathbf{V} = \text{diag}(c_1, \ldots, c_a) + \mathbf{\Sigma})$. Using (4.5.17) and the fact that $\widehat{\mathbf{V}}$ is a consistent estimator of $\mathbf{V}$, we obtain the result of Theorem 4.3.6 for random covariates.

The result for the nonrandom covariate case, follows from the fact that $\widehat{G}$ replaces G in (4.5.17), from (4.5.18), (4.5.20), and the fact that $\widehat{c}_i$ is consistent a estimator of c_i for $i = 1, \ldots, a$.□

4.5.2 Some Useful Lemmas and Auxiliary Results

Lemma 4.5.1. *Let $X_{i1}, \ldots, X_{in_i}$ be order statistics of the covariate from group i, then $\forall\ 0 < \delta < 1$, and $\forall\ j = 1, \ldots, n_i - 1$, $X_{i,j+1} - X_{ij} = o\left(n_i^{-(1-\delta)}\right)$, almost surely.*

Proof. Set $S_j = G_i(X_{i,j+1}) - G_i(X_{ij})$, $j = 1, \ldots, n_i - 1$, where $G_i(\cdot)$ is the distribution function of the covariate in group i. Thus, the S_j are spacings of a sample of size n_i from $U(0,1)$. By the result of Pyke (1965), the distribution function of S_j is

$$F_{S_j}(u) = 1 - (1-u)^{n_i}, \quad \forall\ j. \tag{4.5.25}$$

To show $X_{i,j+1} - X_{ij} = o\left(n_i^{-(1-\delta)}\right)$, almost surely, we will first show that $S_j = o\left(n_i^{-(1-\delta)}\right)$ almost surely. Indeed, for any $\epsilon > 0$, we have

$$\sum_{n_i=1}^{\infty} P\left(\left|\frac{S_j}{n_i^{-(1-\delta)}}\right| > \epsilon\right) = \sum_{n_i=1}^{\infty} P\left(S_j > \epsilon n_i^{-(1-\delta)}\right) = \sum_{n_i=1}^{\infty}\left(1 - \frac{\epsilon}{n_i^{1-\delta}}\right)^{n_i} < \infty,$$

since $\left(1 - \epsilon/n_i^{1-\delta}\right)^{n_i} \sim e^{-\epsilon n_i^{\delta}}$. By the mean value theorem, $S_j = g_i(X_{ij}^*)(X_{i,j+1} - X_{i,j})$ for some $X_{ij}^* \in (X_{ij}, X_{i,j+1})$. Since $g_i(\cdot)$ is uniformly bounded away from 0, our conclusion follows. □

Lemma 4.5.2. *Let $X_{11}, \ldots, X_{1n_1}$ be a random sample from a population with density $g_1(x)$ and distribution function $G_1(x)$, $X_{21}, \ldots, X_{2n_2}$ be a random sample from a population with density $g_2(x)$ and distribution function $G_2(x)$. Assume that the two samples are independent, $n_1/(n_1+n_2) \to \lambda \in (0,1)$, and $g_1(x)$, $g_2(x)$ are differentiable, with common bounded support. For an arbitrary element X_{1j} from the first sample,*

$$N_j = \sum_{j'=1}^{n_2} I(X_{1j} \in W_{1,X_{2j'}}) \leq C_1 n_2^{\delta},$$

$\forall\ 0 < \delta < 1$, for some positive constant C_1 almost surely, for n_2 large enough, where $W_{1,X_{2j'}}$ is the nearest neighborhood window of size k formed around center $X_{2j'}$ using observations from the first sample.

Proof. Letting $A_{n_2} = [N_j \leq C_1 n_2^{\delta}]$, we need to show that

$$P(\{A_{n_2}, \text{ for all } n_2 \text{ except a finite number}) = P(\cup_{n_2=1}^{\infty} \cap_{k=n_2}^{\infty} A_k) = 1,$$

or, equivalently that $P(\cap_{n_2=1}^{\infty} \cup_{k=n_2}^{\infty} A_k^c) = 0$. By the Borel-Cantelli Lemma, it is sufficient to show $\sum_{n_2=1}^{\infty} P(A_{n_2}^c) < \infty$. Note that with $p = (k-1)/2$,

$$I(X_{1j} \in W_{1,X_{2j'}}) = I(X_{2j'} \in [X_{1,j-p}, X_{1,j+p}]). \tag{4.5.26}$$

Thus, $N_j = \sum_{j'=1}^{n_2} I(X_{2j'} \in [X_{1,j-p}, X_{1,j+p}]) \sim Binomial(n_2, P_2)$, where

$$P_2 = G_2(X_{1,j+p}) - G_2(X_{1,j-p}) = g_2(X_{1,j}^*)(X_{1,j+p} - X_{1,j-p}) = o(n_1^{-(1-\delta)}),$$

almost surely, by Lemma 4.5.1. Note that the above relation implies the existence of a constant $C_1 > 0$ such that $C_1 n_2^\delta > n_2 P_2 + n_2^\delta$. Hence

$$\begin{aligned} P(A_{n_2}^c) &= P(N_j > C_1 n_2^\delta) \leq P(N_j - n_2 P_2 > n_2^\delta) \\ \leq & \ P\left(|N_j - n_2 P_2| > n_2^\delta\right) \leq 2exp\left(-\frac{n_2^{2\delta}}{2P_2(1-P_2) + \frac{2}{3} n_2^\delta}\right) \leq 2exp\left(-\frac{n_2^\delta}{2C_1 + \frac{2}{3}}\right), \end{aligned}$$

where the second inequality is an application of Bernstein's inequality for the binomial case. Since $\sum_{n_2=1}^{\infty} exp\left(-\frac{n_2^\delta}{2C_1+\frac{2}{3}}\right) < \infty$, the lemma follows. □

Lemma 4.5.3. *Consider pairs of observations* (X_i, Y_i), $i = 1, \ldots, n$, *and set* $\mathbf{X} = (X_1, \ldots, X_n)'$, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$. *Let* $H_1(\mathbf{X}, \mathbf{Y})$ *and* $H_2(\mathbf{X})$ *be statistics such that*

$$H_1(\mathbf{X}, \mathbf{Y})|\mathbf{X} \to N_{a_1}(0, \mathbf{\Sigma}_1), \quad H_2(\mathbf{X}) \to N_{a_2}(0, \mathbf{\Sigma}_2),$$

where $\mathbf{\Sigma}_1$ *and* $\mathbf{\Sigma}_2$ *are constant nonnegative definitive covariance matrices. Then*

$$\begin{pmatrix} H_1(\mathbf{X}, \mathbf{Y}) \\ H_2(\mathbf{X}) \end{pmatrix} \to N_{a_1+a_2}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{pmatrix}\right).$$

Proof. For any $t_1 \in \mathbf{R}^{a_1}$ and $t_2 \in \mathbf{R}^{a_2}$, we have

$$\begin{aligned} P(H_1(\mathbf{X}, \mathbf{Y}) \leq t_1, H_2(\mathbf{X}) \leq t_2) &= EP(H_1(\mathbf{X}, \mathbf{Y}) \leq t_1, H_2(\mathbf{X}) \leq t_2|\mathbf{X}) \\ &= E\left[P(H_1(\mathbf{X}, \mathbf{Y}) \leq t_1|\mathbf{X}) I(H_2(\mathbf{X}) \leq t_2)\right], \end{aligned}$$

since

$$P(H_1(\mathbf{X}, \mathbf{Y}) \leq t_1, H_2(\mathbf{X}) \leq t_2|\mathbf{X}) = \begin{cases} 0, & \text{if } H_2(\mathbf{X}) > t_2, \\ P(H_1(\mathbf{X}, \mathbf{Y}) \leq t_1|\mathbf{X}), & \text{if } H_2(\mathbf{X}) \leq t_2. \end{cases}$$

Let $\mathbf{Z}_1 \sim N_{a_1}(0, \mathbf{\Sigma}_1)$, then we have

$$\begin{aligned} & E\left[P(H_1(\mathbf{X}, \mathbf{Y}) \le t_1|\mathbf{X})I(H_2(\mathbf{X}) \le t_2) - P(\mathbf{Z}_1 \le t_1)I(H_2(\mathbf{X}) \le t_2)\right] \\ = \;& E[(P(H_1(\mathbf{X}, \mathbf{Y}) \le t_1|\mathbf{X}) - P(\mathbf{Z}_1 \le t_1))I(H_2(\mathbf{X}) \le t_2)] \to 0 \end{aligned}$$

by the dominated convergence theorem, and the assumed convergence in conditional distribution. Thus

$$\begin{aligned} & \lim P(H_1(\mathbf{X}, \mathbf{Y}) \le t_1, H_2(\mathbf{X}) \le t_2) = \lim\{P(\mathbf{Z}_1 \le t_1)E\left[I(H_2(\mathbf{X}) \le t_2)\right]\} \\ = \;& P(\mathbf{Z}_1 \le t_1) \lim P(H_2(\mathbf{X}) \le t_2) = P(\mathbf{Z}_1 \le t_1)P(\mathbf{Z}_2 \le t_2), \end{aligned}$$

where $\mathbf{Z}_2 \sim N_{a_2}(0, \mathbf{\Sigma}_2)$, and hence the result follows. □

Proof of (4.5.3)(ii)

By Chebyshev's inequality it suffices to show $Var(\xi_i) \to 0$. This will follow by showing

$$(i)\ Var(\xi_i|\mathbf{X}_i) \to 0 \ \text{ and } \ (ii)\ E(\xi_i|\mathbf{X}_i) \to c_i. \tag{4.5.27}$$

Indeed, (4.5.27) implies $E(\xi_i^2|\mathbf{X}_i) \to c_i^2$, and thus by the dominated convergence theorem we have $E(\xi_i^2) \to c_i^2$. Combining this with (4.5.3)(i), $Var(\xi_i) \to 0$ follows. Write

$$\begin{aligned} Var(\xi_i|\mathbf{X}_i) \;=\; & \frac{1}{N^2k^4}\sum_{j=1}^{n_i} Var\left\{\sigma_i^2(X_{ij})\left[\sum_{r=1}^{N} I(j \in W_{ir})\right]^2 |\mathbf{X}_i\right\} \\ & + \frac{1}{N^2k^4}\sum_{j_1 \ne j_2}^{n_i} Cov\left\{\sigma_i^2(X_{ij_1})\left[\sum_{r=1}^{N} I(j_1 \in W_{ir})\right]^2, \right. \\ & \left. \sigma_i^2(X_{ij_2})\left[\sum_{r=1}^{N} I(j_2 \in W_{ir})\right]^2 |\mathbf{X}_i\right\}. \end{aligned} \tag{4.5.28}$$

We will show first that the variances and covariances appearing in (4.5.28) are finite. This will follow from the finiteness of $E\left\{\sigma_i^4(X_{ij})\left[\sum_{r=1}^{N} I(j \in W_{ir})\right]^4 |\mathbf{X}_i\right\}$, which we now show. By assumption A3 it suffices to consider the expectation of

$$\begin{aligned}\left[\sum_{r=1}^{N} I(j \in W_{ir})\right]^4 &= \sum_{r=1}^{N} I(j \in W_{ir}) + \binom{4}{2} \sum_{r_1 \neq r_2} I(j \in W_{ir_1}) I(j \in W_{ir_2}) \\ &+ \binom{4}{3} \sum_{\substack{r_1,\ r_2,\ r_3 \\ are\ distinct}} I(j \in W_{ir_1}) I(j \in W_{ir_2}) I(j \in W_{ir_3}) \\ &+ \sum_{\substack{r_1, r_2, r_3, r_4 \\ are\ distinct}} I(j \in W_{ir_1}) I(j \in W_{ir_2}) I(j \in W_{ir_3}) I(j \in W_{ir_4}).\end{aligned}$$

According to the notation in (4.5.6), $E\left(\sum_{r=1}^{N} I(j \in W_{ir}) | \mathbf{X}_i\right) = Q_{11,j}(\mathbf{X}_i) + Q_{12,j}(\mathbf{X}_i)$ and $E\left(\sum_{r_1 \neq r_2}^{N} I(j \in W_{ir_1} \cap W_{ir_2}) | \mathbf{X}_i\right) = Q_{21,j}(\mathbf{X}_i) + Q_{22,j}(\mathbf{X}_i) + Q_{23,j}(\mathbf{X}_i)$, whose expected values were shown there to be finite. The finiteness of the expectations of the other two terms can be seen exactly the same way. Thus, the variances and covariances in (4.5.28) are finite, which also shows that the first term in (4.5.28) tends to zero. To show that the second term of (4.5.28) tends to zero, we first note that it suffices to replace the double sum $\sum_{j_1 \neq j_2}$ in this term by $\sum_{|j_1 - j_2| > k}$. Also note that by assumption A3,

$$\begin{aligned}&\left| Cov\left\{ \sigma_i^2(X_{ij_1}) \left[\sum_{r=1}^{N} I(j_1 \in W_{ir})\right]^2, \sigma_i^2(X_{ij_2}) \left[\sum_{r=1}^{N} I(j_2 \in W_{ir})\right]^2 | \mathbf{X}_i \right\}\right| \\ \leq\ & C_4 \left| Cov\left\{ \left[\sum_{r=1}^{N} I(j_1 \in W_{ir})\right]^2, \left[\sum_{r=1}^{N} I(j_2 \in W_{ir})\right]^2 | \mathbf{X}_i \right\}\right|,\end{aligned}$$

for some positive constant C_4. Thus, we must evaluate $Cov(\Gamma_i^2(j_1), \Gamma_i^2(j_2) | \mathbf{X}_i)$, where

$$\Gamma_i(j_1) = \sum_{r=1}^{N} I(j_1 \in W_{ir}) = k + \sum_{i' \neq i}^{a} \Gamma_{ii'}(j_1),$$

$$\Gamma_i(j_2) = \sum_{r=1}^{N} I(j_2 \in W_{ir}) = k + \sum_{i' \neq i}^{a} \Gamma_{ii'}(j_2),$$

with $\Gamma_{ii'}(j_1) = \sum_{j=1}^{n_{i'}} I(j_1 \in W_{i,X_{i'j}})$, $\Gamma_{ii'}(j_2) = \sum_{j=1}^{n_{i'}} I(j_2 \in W_{i,X_{i'j}})$. Since $\Gamma_{ii'}(j_1)$ and $\Gamma_{ii'}(j_2)$ are conditionally independent given $\mathbf{X}_i$ if $i'_1 \neq i'_2$,

$Cov(\Gamma_i^2(j_1), \Gamma_i^2(j_2)|\mathbf{X}_i)$ is a finite sum of terms like $Cov(\Gamma_{ii'}(j_1), \Gamma_{ii'}(j_2)|\mathbf{X}_i)$, $Cov(\Gamma_{ii'}(j_1)^2, \Gamma_{ii'}(j_2)|\mathbf{X}_i)$, $Cov(\Gamma_{ii'}(j_1), \Gamma_{ii'}(j_2)^2|\mathbf{X}_i)$, $Cov(\Gamma_{ii'}(j_1)^2, \Gamma_{ii'}(j_2)^2|\mathbf{X}_i)$, and $Cov(\Gamma_{ii'_1}(j_1)\Gamma_{ii'_2}(j_1), \Gamma_{ii'_1}(j_2)\Gamma_{ii'_2}(j_2)|\mathbf{X}_i)$.

Because we are considering only the case that $|j_1 - j_2| > k$, it follows that the set of windows containing X_{ij_1} is distinct from that containing X_{ij_2}. Thus, if we define $\Gamma_{ii'}(j_1, j_2) = \sum_{j=1}^{n_{i'}} I(j_1 \notin W_{i,X_{i'j}}, j_2 \notin W_{i,X_{i'j}})$, the vector $(\Gamma_{ii'}(j_1), \Gamma_{ii'}(j_2), \Gamma_{ii'}(j_1, j_2))$ has the trinomial distribution. According to Lemma 4.5.1, the probabilities for each of the three outcomes are $\pi_1 = o(n_{i'}^{-(1-\delta)})$, $\pi_2 = o(n_{i'}^{-(1-\delta)})$ and $\pi_3 = 1 - \pi_1 - \pi_2$. With this observation all of the above covariances can be obtained from the trinomial moment generating function $M(\mathbf{t}) = (\pi_1 e^{t_1} + \pi_2 e^{t_2} + (1 - \pi_1 - \pi_2)e^{t_3})^{n_{i'}}$. For example,

$$\begin{aligned} Cov(\Gamma_{ii'}^2(j_1), \Gamma_{ii'}^2(j_2)|\mathbf{X}_i) &= \frac{\partial M(\mathbf{t})}{\partial t_1^2 t_2^2}|_{t_1=t_2=t_3=0} - \left(\frac{\partial M(\mathbf{t})}{\partial t_1^2}|_{t_1=t_2=t_3=0}\right) \\ &\quad \times \left(\frac{\partial M(\mathbf{t})}{\partial t_2^2}|_{t_1=t_2=t_3=0}\right) \\ &= O(n_{i'}\pi_1\pi_2) = O(n_{i'}^{-(1-2\delta)}). \end{aligned}$$

It follows that $N^{-2}\sum_{|j_1-j_2|>k} Cov(\Gamma_i^2(j_1), \Gamma_i^2(j_2)|\mathbf{X}_i) = o(1)$, completing the proof of (4.5.27)(i). Now we proceed with the proof of (4.5.27)(ii). Noting that

$$E(\xi_i|\mathbf{X}_i) = \frac{1}{Nk^2}\sum_{j=1}^{n_i} \sigma_i^2(X_{ij})\left[Q_{11,j}(\mathbf{X}_i) + Q_{12,j}(\mathbf{X}_i) + Q_{21,j}(\mathbf{X}_i) + Q_{22,j}(\mathbf{X}_i) + Q_{23,j}(\mathbf{X}_i)\right],$$

it suffices to show

$$\frac{1}{Nk^2}\sum_{j=1}^{n_i} \sigma_i^2(X_{ij})Q_{m,j}(\mathbf{X}_i) = c_{i,m} + o(1), \quad \forall\, m \in \mathcal{I} = \{11, 12, 21, 22, 23\}, \tag{4.5.29}$$

where

$$\begin{aligned}
c_{i,11} &= \frac{c_{i,21}}{k-1} = \frac{\lambda_i}{k} E\left\{\sigma_i^2(X_{ij})\right\}, \\
c_{i,12} &= \frac{c_{i,22}}{2k} = \frac{k-1}{k^2} \lambda_i E\left\{\sigma_i^2(X_{ij}) \frac{g(X_{ij}) - \lambda_i g_i(X_{ij})}{g_i(X_{ij})}\right\}, \\
c_{i,23} &= \frac{k(k-1)}{k^2} E\left\{\sigma_i^2(X_{ij}) \frac{\left[g(X_{ij}) - \lambda_i g_i(X_{ij})\right]^2}{\lambda_i g(X_{ij})}\right\},
\end{aligned}$$

because $\sum_{m\in\mathcal{I}} c_{i,m} = c_i$. However, by the proof of relation (4.5.3)(i), we already have that $(Nk^2)^{-1} E\left\{\sum_{j=1}^{n_i} \sigma_i^2(X_{ij}) Q_{m,j}(\mathbf{X}_i)\right\} = n_i (Nk^2)^{-1} E\left\{\sigma_i^2(X_{ij}) Q_{m,j}(\mathbf{X}_i)\right\} = c_{i,m} + o(1),\ \ \forall\, m \in \mathcal{I} = \{11, 12, 21, 22, 23\}$. Thus, (4.5.29) will follow if we show that the variance of $(Nk^2)^{-1} \sum_{j=1}^{n_i} \sigma_i^2(X_{ij}) Q_{m,j}(\mathbf{X}_i)$ tends to zero. In view of its aforementioned expectation, this is equivalent to

$$E\left\{\left[\frac{1}{Nk^2} \sum_{j=1}^{n_i} \sigma_i^2(X_{ij}) Q_{m,j}(\mathbf{X}_i)\right]^2\right\} = c_{i,m}^2 + o(1),\ \ \forall\, m \in \mathcal{I}. \qquad (4.5.30)$$

Noting that,

$$\begin{aligned}
&\frac{1}{N^2} E\left\{\left[\sum_{j=1}^{n_i} \sigma_i^2(X_{ij})\right]^2\right\} = \frac{1}{N^2} E\left\{\left[\sum_{j=1}^{n_i} \sigma_i^4(X_{ij}) + \sum_{j_1\neq j_2} \sigma_i^2(X_{ij_1}) \sigma_i^2(X_{ij_2})\right]\right\} \\
&= \frac{1}{N^2}\left\{n_i E\left[\sigma_i^4(X_{ij})\right] + n_i(n_i - 1)\left\{E\left[\sigma_i^2(X_{ij})\right]\right\}^2\right\} \\
&= \lambda_i^2 \left\{E\left[\sigma_i^2(X_{ij})\right]\right\}^2 + o(1),
\end{aligned}$$

and using the fact that $Q_{11,j} = k$, $Q_{21,j} = k(k-1)$, we have that (4.5.30) is true for $m = 12$ and $m = 21$. Next we consider the case with $m = 12$. We have

$$
\begin{aligned}
& E\left\{\left[\frac{1}{Nk^2}\sum_{j=1}^{n_i}\sigma_i^2(X_{ij})Q_{12,j}(\mathbf{X}_i)\right]^2\right\} \\
= \quad & \frac{1}{N^2k^4}E\left[\sum_{j_1\neq j_2}\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})Q_{12,j_1}(\mathbf{X}_i)Q_{12,j_2}(\mathbf{X}_i)\right]+o(1) \\
= \quad & \frac{1}{N^2k^4}\sum_{j_1\neq j_2}\sum_{i_1,i_2\neq i}n_{i_1}n_{i_2}E\left[\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_2})\right. \\
& \left.\times\frac{D_{ij_1}(k-1)D_{ij_2}(k-1)}{g_i(X_{ij_1})g_i(X_{ij_2})}\right]+o(1) \\
= \quad & \frac{1}{N^2k^4}\sum_{j_1\neq j_2}\sum_{i_1,i_2\neq i}n_{i_1}n_{i_2}E\left[\frac{\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_2})}{g_i(X_{ij_1})g_i(X_{ij_2})}\right. \\
& \left.\times E\left\{D_{ij_1}(k-1)D_{ij_2}(k-1)|X_{ij_1},X_{ij_2}\right\}\right]+o(1), \qquad (4.5.31)
\end{aligned}
$$

where $D_{ij}(k-1) = G_i(X_{ij}+U_{ij})-G_i(X_{ij}-L_{ij})$. The first equality is obtained by the fact that $E\{\sigma^4(X_{ij})Q_{j,12}(\mathbf{X}_i)^2\}$ is bounded, which follows from assumption A3 and (4.5.12). The second equality is obtained using (4.5.7). To evaluate $E\{D_{ij_1}(k-1)D_{ij_2}(k-1)|X_{ij_1},X_{ij_2}\}$ we note that in the case $|j_1-j_2|\geq k-1$, $D_{ij_1}(k-1)$ and $D_{ij_2}(k-1)$ have no common spacings, and in the case $|j_1-j_2| < k-1$, $D_{ij_1}(k-1)$ and $D_{ij_2}(k-1)$ have $k-1-|j_1-j_2|$ common spacings. Applying a result of Pyke (1965) we have

$$
\begin{aligned}
E\left\{D_{ij_1}(k-1)D_{ij_2}(k-1)|X_{ij_1},X_{ij_2}\right\} \quad = \quad & \frac{(k-1)^2}{(n_i+1)(n_i+2)} \\
+ \quad & \frac{\min\{(k-1-|j_1-j_2|),\ 0\}}{(n_i+1)(n_i+2)}. \qquad (4.5.32)
\end{aligned}
$$

Since the number of $j_1,\ j_2$ pairs which differ by less than k is $O(n_i)$, it suffices to replace $\sum_{j_1\neq j_2}$, in expression (4.5.31), by $\sum_{|j_1-j_2|\geq k}$. With this substitution, and using (4.5.32), (4.5.31) equals

$$\frac{(k-1)^2}{N^2k^4}\sum_{i_1,i_2\neq i}\frac{n_{i_1}n_{i_2}}{n_i^2}\sum_{|j_1-j_2|\geq k}E\left[\frac{\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_2})}{g_i(X_{ij_1})g_i(X_{ij_2})}\right]+o(1)$$
$$=\frac{(k-1)^2}{N^2k^4}\sum_{i_1,i_2\neq i}n_{i_1}n_{i_2}E\left[\frac{\sigma_i^2(X_{ij_1})g_{i_1}(X_{ij_1})}{g_i(X_{ij_1})}\right]E\left[\frac{\sigma_i^2(X_{ij_2})g_{i_2}(X_{ij_2})}{g_i(X_{ij_2})}\right]+o(1)$$
$$=\frac{(k-1)^2}{k^4}\left\{E\left[\sigma_i^2(X_{ij})\frac{g(X_{ij})-\lambda_i g_i(X_{ij})}{g_i(X_{ij})}\right]\right\}^2+o(1)=c_{i,12}^2+o(1).$$

Next, we can obtain (4.5.30) for $m = 22$, using the fact that $Q_{j,22}(\mathbf{X}_i) = 2kQ_{j,12}(\mathbf{X}_i)$.

It remains to show (4.5.30) for $m = 23$. In order to do so we need to need formulas for the expectations of uniform spacings raised to a power greater than 2, which are not contained in Pyke (1965). Define the sequence of uniform spacings $S_1,\ldots,S_n$ such that $S_1 = U_1$, $S_i = U_i - U_{i-1}$ for $2 \leq i \leq (n-1)$ and $S_n = 1 - S_{n-1}$, where $U_1,\ldots,U_n$ are the order statistics of n independent uniform random variables on $(0,1)$. It can be shown that

$$E(S_i^a) = E(S_1^a) = \binom{n+a}{a}^{-1} \tag{4.5.33}$$
$$E(S_1^2S_2^2) = \frac{2}{n(n+1)}E\left[S_1^2(1-S_1)^2\right] \tag{4.5.34}$$
$$E(S_1^2S_2S_3) = \frac{1}{n(n-1)}E\left[S_1^2(1-S_1)^2\right]-\frac{1}{n-1}E(S_1^2S_2^2) \tag{4.5.35}$$
$$E(S_1S_2S_3S_4) = \frac{E\left[S_1(1-S_1)(3S_1^2-5S_1+1)\right]}{n(n-1)(n-2)}-\frac{E(S_1^2S_2^2)}{(n-1)(n-2)}. \tag{4.5.36}$$

Now Consider the term in (4.5.30) with $m = 23$

$$
\begin{aligned}
& E\left\{\left[\frac{1}{Nk^2}\sum_{j=1}^{n_i}\sigma_i^2(X_{ij})Q_{23,j}(\mathbf{X}_i)\right]^2\right\} \\
= \; & E\left\{\left[\frac{1}{Nk^2}\sum_{j=1}^{n_i}\sigma_i^2(X_{ij})Q_{12,j}(\mathbf{X}_i)^2\right]^2\right\} + o(1) \\
= \; & \frac{1}{N^2k^4}E\left\{\sum_{j_1\neq j_2}\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})Q_{12,j_1}(\mathbf{X}_i)^2 Q_{12,j_2}(\mathbf{X}_i)^2\right\} + o(1) \\
= \; & \frac{1}{N^2k^4}E\left\{\sum_{j_1\neq j_2}\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})\left[\sum_{i_1,i_2\neq i} n_{i_1}n_{i_2}\left(g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_1})\frac{D_{ij_1}(k-1)^2}{g_i(X_{ij_1})^2}\right)\right]\right. \\
& \left.\times\left[\sum_{i_3,i_4\neq i} n_{i_3}n_{i_4}\left(g_{i_3}(X_{ij_2})g_{i_4}(X_{ij_2})\frac{D_{ij_2}(k-1)^2}{g_i(X_{ij_2})^2}\right)\right]\right\} + o(1) \\
= \; & \frac{1}{N^2k^4}\sum_{\substack{i_1,i_2\neq i\\ i_3,i_4\neq i}} n_{i_1}n_{i_2}n_{i_3}n_{i_4}\sum_{j_1\neq j_2} E\left\{\sigma_i^2(X_{ij_1})\sigma_i^2(X_{ij_2})\frac{g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_1})}{g_i(X_{ij_1})^2}\right. \\
& \left.\times\frac{g_{i_3}(X_{ij_2})g_{i_4}(X_{ij_2})}{g_i(X_{ij_2})^2}E\left[D_{ij_1}(k-1)^2 D_{ij_2}(k-1)^2 | X_{ij_1}, X_{ij_2}\right]\right\} + o(1). \qquad (4.5.37)
\end{aligned}
$$

The first equality follows by (4.5.10). The second equality is justified since we have shown that $E[\sigma_i^2(X_{ij_1})Q_{12,j}(\mathbf{X}_i)^2]$ is bounded, see (4.5.9), and thus $E[\sigma_i^4(X_{ij_1})Q_{12,j}(\mathbf{X}_i)^4]$ is bounded. In the cases where $|j_1 - j_2| \geq k-1$, using (4.5.33)-(4.5.36), we have

$$
\begin{aligned}
& E\left[D_{ij_1}(k-1)^2 D_{ij_2}(k-1)^2 | X_{ij_1}, X_{ij_2}\right] \\
& = (k-1)^2E(S_1^2S_2^2) + 2(k-1)^2(k-2)E(S_1^2S_2S_3) + (k-1)^2(k-2)^2E(S_1S_2S_3S_4) \\
& = \frac{2(k-1)^2}{n_i(n_i-1)}E(S_1^2) + \frac{2(k-1)^2(k-2)}{n_i(n_i-1)}E(S_1^2) + \frac{(k-1)^2(k-2)^2}{n_i(n_i-1)(n_i-2)}E(S_1) + o(n_i^{-4}) \\
& = \frac{k^2(k-1)^2}{n_i^4} + o(n_i^{-4}). \qquad (4.5.38)
\end{aligned}
$$

Since the number of pairs (j_1, j_2) which differ by less than k is $O(n_i)$, and $E\left[D_{ij_1}(k-1)^2 \;\; D_{ij_2}(k-1)^2 | X_{ij_1}, X_{ij_2}\right] = O(n_i^{-4})$, it suffices to replace $\sum_{j_1\neq j_2}$

in (4.5.37), by $\sum_{|j_1-j_2|\geq k}$. With this substitution, and using (4.5.38), (4.5.37) becomes

$$\frac{k^2(k-1)^2}{n_i^4N^2k^4}\sum_{\substack{i_1,i_2\neq i\\ i_3,i_4\neq i}} n_{i_1}n_{i_2}n_{i_3}n_{i_4}\sum_{|j_1-j_2|\geq k} E\left\{\sigma_i^2(X_{ij_1})\frac{g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_1})}{g_i(X_{ij_1})^2}\right.$$

$$\left.\times\sigma_i^2(X_{ij_2})\frac{g_{i_3}(X_{ij_2})g_{i_4}(X_{ij_2})}{g_i(X_{ij_2})^2}\right\}+o(1)$$

$$=\frac{n_i^2k^2(k-1)^2}{n_i^4N^2k^4}\left\{E\left[\sigma_i^2(X_{ij_1})\sum_{i_1,i_2\neq i}\frac{n_{i_1}n_{i_2}g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_1})}{g_i(X_{ij_1})^2}\right]\right\}^2+o(1)$$

$$=\frac{k^2(k-1)^2}{k^4}\left\{E\left[\sigma_i^2(X_{ij_1})\sum_{i_1,i_2\neq i}\frac{\lambda_{i_1}\lambda_{i_2}g_{i_1}(X_{ij_1})g_{i_2}(X_{ij_1})}{\lambda_i g_i(X_{ij_1})^2}\right]\right\}^2+o(1)$$

$$=c_{i,23}^2+o(1).$$

This completes the proof. □

Chapter 5

Summary and Future Research

5.1 Summary and Conclusions

In conclusion, this thesis provides new approaches to fully nonparametric testing procedures in the cases of factorial designs with repeated measurements MCR and in ANCOVA. The proposed methodologies are valid for ordinal or continuous data and the hypotheses we consider are functional nonparametric hypotheses, which have the advantage of being invariant under monotone transformation of the data and do not depend on any modelling assumptions

The first part of the thesis, Chapters 2 and 3, provides a new perspective of dealing with missing data problems with a fully nonparametric approach. First we extend the simple matched pairs analysis proposed by Akritas et. al. (2002) to factorial designs with paired observations and then we generalize this method to factorial designs with more than two repeated measurements. This is the first time (mid-) rank testing procedures are applied to missing data without assuming the data are MCAR. The methodology also includes a contribution in the literature of nonparametric imputation by relaxing the assumption of strong ignorability. This is achieved though a representation of the marginal distributions, which allowed the use of more general and flexible donor sets in our nonparametric imputation. The donor set selection basically determines the assumptions regarding the missingness pattern. In the paired data cases the choice of donor set is straightforward but in cases with more than two repeated measurements this choice can vary depending on the nature of the data set, sample size limitations e.t.c. We investigated the performance of these tests in terms of achieved Type I error and power using different simulation schemes comparing it with ML method and the nonparametric complete cases method. We also present a real data application, both for a case of paired data and for a case of four repeated measurements.

The second part of the thesis, Chapter 4, provides nonparametric methods for ANCOVA. We consider the fully nonparametic ANCOVA model proposed by Akritas, Arnold and Du (2000) which avoids the strong assumptions required for the classical ANCOVA, such as the restrictive assumption of normality, homoscedasticity and linearity. Their approach relies on consistent estimation (using kernel methods) of the conditional distribution functions F_{ix}. By its nature, application of this approach requires determination of the window bandwidth which is particularly cumbersome in this context. (In Akritas, Arnold and Du (2000) the bandwidth choice was based on resampling from the data in a way that imitates the null hypothesis.) Here we considered an alternative test procedure for the same hypotheses. The main novelty of the new procedure is that it does not require consistent estimation of the F_{ix} and the estimator is obtained using nearest neighbor windows of *fixed* size. The basic idea for the proposed testing procedure is to think of the covariate as another factor with infinitely many levels, and consider test statistics used in two way ANOVA when one of the factors has many levels; see Wang and Akritas (2002). Since there is only one observation per covariate value, this hypothetical two-way ANOVA design has at most one observation per cell. To remedy this, we use smoothness assumptions and augment the cells with observations falling in a window around each covariate value. We remark that the artificial ANOVA design will have dependent observations, since each observation will belong to several groups, but this creates only a minor theoretical problem which is overcome by the use of a suitable central limit theorem for dependent observations. The asymptotic distribution of the test statistic is derived under both the null and the alternative hypotheses, with the covariate being either random or non-random. Simulation results indicate that the model is effective in term of achieved Type I error and power, both for random and non-random covariates.

5.2 Future Research

Fully nonparametric testing procedures for missing data is a new research area and there are many issues that could be addressed in future research. One important direction is to consider designs with a large number of repeated measurements. Thus one has a response curve for each subject, and the resulting data are also called functional data. The condition of identical observation time points for all subjects is unrealistic in such data sets. One way of dealing with not identical time points is to consider the set of all time points as common for all subjects and treat the resulting slots (time points) as missing observations. This is one case where the assumption of MCAR is realistic. Of course we can have truly missing observations wither in designs with identical time points or not. For this type of designs nonparametric imputation can use not only between subjects information (donor sets) but also within subject information using individual response curve. We note that Wang and Akritas (2004) consider the case of a large number of measurements per subject at identical time points with complete observations. Their methodology uses a CLT for α-mixing processes and thus it relies on a large number if observations per subject instead of a large number of subjects.

Another interesting direction is to investigate designs where there are a few repeated measurements subject to missingness and a large number of factor levels. This designs have applications in cases where there are a few factors of primary interest and a lot of factors which are not of primary interest. Factors not of primary interest may include auxiliary factors making the MCR assumption more realistic. To give a brief description of the basic idea consider the simpler case with paired data, the first time point fully observed and collapse all auxiliary factors into one factor with large number of levels (so the only factor of primary interest is time). Thus for each subject we have $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})$, with $Y_{ijk} \sim F_{ij}$, for $i = 1, \ldots, I$, $j = 1, 2$ and $k = 1, \ldots, n_i$. To test the time effect we will need to consistently estimate the $\int H(y) d\ \bar{F}_{.j}$, where $\bar{F}_{.j}(y) = 1/I \sum_i F_{ij}(y)$, for $j = 1, 2$. With some asymptotic results this problem reduces to the one of consistently estimating the marginal distribution functions $\bar{F}_{.j}$. For the first time point, where the data is fully observed, $\bar{F}_{.1}$ is consistently estimated by the average of the empirical distribution functions $\widehat{F}_{i1}$, $i = 1, \ldots, I$. The challenge in this problem is to achieve consistent estimation of $\bar{F}_{.2}$. By a decomposition similar to that used before, under MCR we have

$$
\begin{aligned}
F_{i2}(y) &= P(\Delta_{i2} = 1)\ F_{i2}(y|\Delta_{i2} = 1) \\
&+ P(\Delta_{i2} = 0) \int F_{i2}(y|Y_{i1} = x, \Delta_{i2} = 1) d\ F_{i1}(x|\Delta_{i2} = 0),
\end{aligned}
$$

Because individual estimates of each $F_i 2$ will be averaged to produce consistent estimate of $\bar{F}_{.2}$, the individual estimates need not be consistent. However, they need to be unbiased (or the bias to tend to zero suitably fast). Since we condition on the Y_{i1} values with $\Delta_{i2} = 0$, we cannot have unbiased estimator of $F_{i2}(y|Y_{i1} = x, \Delta_{i2} = 1)$ even with window containing only one observation. The challenge is to find the lowest rate at which the $n_i \to \infty$ resulting in consistent estimator of $\bar{F}_{\cdot 2}$.

Fully nonparametric analysis of covariance is a relatively new research area which is very important in a lot of applications. In this thesis we made a step towards the further development of this model based on new methodologies. The next step to be considered is to apply the proposed procedure to the (mid-) rank transformed data rather than on the original observations. This will make out methodology more attractive and more robust to extreme data behavior.

Some other designs with a lot of applications are ANCOVA with dependent data and ANCOVA with more than one covariates. The first extension, to dependent data, it is a rather simple step from the simple ANCOVA design we presented in Chapter 4. The proposed method is promising for extensions to designs to more than one covariates without coming across with the "curse of dimensionality".

Moreover we plan to apply the ANCOVA methodology to missing data. Thus we will considering the covariate with many levels and extend the previously described results to this case. Preliminary investigations suggest that smoothing assumption (which are not appropriate in the previously described problem with many factors) will allow as to produce consistent estimator of the need nonparametic quantities using windows of fixed size.

L

APPENDIX

Appendix A

R CODES FOR SIMULATIONS

A.1 Code for Generating Missing data

```
##----------------------------------------------------------------------------
## FUNCTION FOR SIMULATING MISSING DATA
## It can handle:
##       - DIFFERENT % MISS. IN EACH CELL (pm.mat)
##       - HETEROSCEDASTISITI (m.mat,v.mat)
##       - DIFFERENT RHO'S IN EACH GROUP (rho.vec)
##----------------------------------------------------------------------------
## ---------------------------------------------------------------------------
## Function to create AR cov matrix.
AR <- function(a,rho){
  Sigma <- matrix(rep(0,a*a),a,a,)
  for(k in 0:(a-1))for(i in 1:(a-k))Sigma[i,i+k] <- Sigma[i+k,i] <- rho^k
  Sigma
}

## pm.mat = probability of missigness in each cell
## m.mat = means in every cell (Default: mu_ij=0 for Normal, =1 for Gamma)
## v.mat = var in every cell (Default: v_ij=1 for Normal, not used for Gamma)
## mp = 1 -> MCAR, mp = 2 -> MAR
## dist =0 -> Normal, dist =1 ->lognormal dist =2 -> Gamma (m>0)

smd <- function(a,N.vec,mp,dist,rho.vec,pm.mat,m.mat,v.mat,tau.m){
  result <- list()
  I <- length(N.vec)
  if(missing(tau.m))tau.m <- matrix(rep(0,(a*I)),I,a)
  if(missing(m.mat))m.mat <- matrix(rep(0,(a*I)),I,a) + 1*(dist==2)
  if(missing(v.mat))v.mat <- matrix(rep(1,(a*I)),I,a)
  m.mat <- m.mat+tau.m/sqrt(sum(N.vec)/I)
  for(i in 1:I){
    m <- m.mat[i,]
```

```
rho <- rho.vec[i]
N <- N.vec[i]
s <- AR(a,rho.vec[i])
diag(s) <- v.mat[i,]
s <-t(chol(s))
if(dist==0) Y <- t(sapply(rep(a,N.vec[i]),function(x)s%*%rnorm(x)+m))
if(dist==1){
  Y <- t(sapply(rep(a,N.vec[i]),function(x)exp(s%*%rnorm(x)+m)))
  Y[(Y  > 10)] <- 10
}
if(dist==2){# for bivariate gamma, m>0
  a0 <- rho.vec[i]*sqrt(prod(m))
  beta <- 1;
  Y0 <- rgamma(N.vec[i], a0)
  Y <- numeric()
  for(j in 1:a){
    aj <- beta*m[j] -a0
    Yj <- rgamma(N.vec[i], aj)
    Y <- cbind(Y, Y0+Yj)
  }
}

## ##################################
## Create the missigness mechanism. #
## ##################################
if(mp==1){ ## MCAR...
  d <- matrix(rep(1,a*N.vec[i]),N.vec[i],a); it <- 0
  pm <- as.matrix(pm.mat[i,])
  ne0 <- sum(pm>0)
  while(nrow(unique(d[!(apply(d,1,sum)==0),])) < 2^ne0-1*(ne0==a)){
    d <- apply(1-pm,1,rbinom,n=N.vec[i],size=1)
    it <- it+1
  }
  Y[d==0] <- NA
  Y <- Y[(apply(d,1,sum)>0),]
  d <- d[(apply(d,1,sum)>0),]
  result[[i]] <- list("Y"=Y,"d"=d)
}
if(mp==2){ ## MAR ...
  d <- matrix(rep(1,a*N.vec[i]),N.vec[i],a); it <- 0
  pm <- as.matrix(pm.mat[i,])
  ne0 <- sum(pm>0)
  while(nrow(unique(d[!(apply(d,1,sum)==0),])) < 2^ne0-1*(ne0==a)){
```

```
      dd <- matrix(rep(1,a*N.vec[i]),N.vec[i],a)
      dd[,1] <- rbinom(n=N.vec[i],size=1,prob=(1-pm[1]))
      b1 <-log(2)
      for(j in 2:a){
        p <- pm[j]
        n0 <- sum(dd[,(j-1)]==0)
        d0 <- rbinom(n=n0,size=1,prob=(1-p))
        y1 <- Y[(dd[,(j-1)]==1),(j-1)]
        b0 <- log(p/(1-p))-b1*mean(y1)
        p1 <- exp(apply(cbind(1,y1),1,function(v)v%*%c(b0,b1)))
        p1 <- as.matrix(p1/(1+p1))
        d1 <- apply(1-p1,1,rbinom,n=1,size=1)
        dd[(dd[,(j-1)]==0),j] <- d0
        dd[(dd[,(j-1)]==1),j] <- d1
        }
      d <- dd
      it <- it+1
    }
    Y[d==0] <- NA
    Y <- Y[(apply(d,1,sum)>0),]
    d <- d[(apply(d,1,sum)>0),]
    result[[i]] <- list("Y"=Y,"d"=d)
  }
  if(mp==21){ ## MAR ...
    d <- matrix(rep(1,a*N.vec[i]),N.vec[i],a); it <- 0
    pm <- as.matrix(pm.mat[i,])
    ne0 <- sum(pm>0)
    while(nrow(unique(d[!(apply(d,1,sum)==0),])) < 2^ne0-1*(ne0==a)){
      dd <- matrix(rep(1,a*N.vec[i]),N.vec[i],a)
      dd[,1] <- rbinom(n=N.vec[i],size=1,prob=(1-pm[1]))

      for(j in 2:a){
        p <- pm[j]
        n0 <- sum(dd[,(j-1)]==0)
        d0 <- rbinom(n=n0,size=1,prob=(1-p))
        y1 <- Y[(dd[,(j-1)]==1),(j-1)]

        p1 <- as.matrix(apply(as.matrix(y1),1,function(v){
          (.6-.2*abs(v))*(abs(v)<3)+.1*(abs(v)>=3)}))
        d1 <- apply(1-p1,1,rbinom,n=1,size=1)
        dd[(dd[,(j-1)]==0),j] <- d0
        dd[(dd[,(j-1)]==1),j] <- d1
        }
```

```
        d <- dd
        it <- it+1
      }
      Y[d==0] <- NA
      Y <- Y[(apply(d,1,sum)>0),]
      d <- d[(apply(d,1,sum)>0),]
      result[[i]] <- list("Y"=Y,"d"=d)
    }
  }
  result
}
```

A.2 Code for Mixed-Effects Models with MCR Data

```
## ############################################################
## SIMULATION STUDY FUNCTIONS
## FACTORIAL DESIGNS
##   - Time Factor ( for j=1,..., a; a>=2)
##   - Group Factor (i = 1,...,I; I>=1)
## ############################################################

## -------------#
## Indices i_r  #
## -------------#
Q2 <- matrix(c(1,2,2,1),2,2)
Q3 <- t(matrix(c(1,2,3,2,1,3,3,2,1),3,3))
Q4 <- matrix(c(1,2,3,4,2,1,2,3,3,3,4,2,4,4,1,1),4,4)

##------------------------------------#
## FUNCTION USED IN CONTARST MATRICES #
##------------------------------------#
M.fun <- function(v){
  if(v==2)result <- t(c(1,-1))
  if(v>2)result <- cbind(rep(1,(v-1)),diag(rep(-1,(v-1))))
result
}

## -----------------#
## Kernel Functions #
## -----------------#
K.normal <- function(x,b)1/b*1/sqrt(2*pi)*exp(-(x/b)^2/2)
```

```
K.epanechnikov <- function(x,b)(3/4)*(1-(x/b)^2*(abs(x/b)<1))

## ---------------------------------------------------------------------#
## Data Arrangement based on missigness patterns S^j_r and tilde(S^j_r) #
## ---------------------------------------------------------------------#
fS <- function(Y,d,i,r,a,Q){
  if(r==1){
    result <- Y[(d[,i]*d[,Q[i,2]]==1),Q[i,]]
    if(a > 2){
      for(r1 in 3:a)
        result <- rbind(result,Y[(apply(cbind(d[,i],(1-d[,Q[i,2:(r1-1)]]),
                                d[,Q[i,r1]]),1,sum)==r1),Q[i,]])
    }
    result <- rbind(result,Y[(apply(cbind(d[,i],
                         (1-d[,Q[i,2:a]])),1,sum)==a),Q[i,]])
    ## to avoid vector form if only one case.
    result <- matrix(as.vector(result),length(result)/a,a)
  }
  else result <- Y[(apply(cbind((1-d[,Q[i,1:(r-1)]]),d[,Q[i,r]]),1,sum)
                             ==r),Q[i,r]]
  as.matrix(result)
}

## Always for r>1!
fs <- function(Y,d,i,r,Q){
  if(r==2)result <- Y[(d[,i]*d[,Q[i,r]]==1),c(i,Q[i,r])]
  if(r>2)result <- Y[(apply(cbind(d[,i],(1-d[,Q[i,2:(r-1)]]),d[,Q[i,r]]),
                         1,sum)==r), c(i,Q[i,r])]
  ## to avoid vector form if only one case.
  result <- matrix(as.vector(result),length(result)/2,2)
  result
}

## -------------------------------------------------------------------------------#
## estimation function:                                                           #
##  - Returns the test statistic vector and Covariance matrix for ONE group.     #
##  - group.data.list[[1]] = Y, data in one group                                 #
##                   [[2]] = d, corresponding d (0 if NA, 1 otherwise)            #
##                   [[3]] = H(Y)                                                 #
##                   [[4]] = b, bandwidth to be used for this group.              #
##  - a = number of time points                                                   #
##  - Q = matrix of j_r indices                                                   #
##  - ir = indices combination list used as a function argument (see Test fun.) #
```

```
## -------------------------------------------------------------------------------#
estimation.function <- function(group.data.list,a,Q,ir,kernel){
  if(kernel==0) K <- K.normal
  if(kernel==1) K <- K.epanechnikov
  Y <- group.data.list[[1]]
  d <- group.data.list[[2]]
  HY <- group.data.list[[3]]
  b <- group.data.list[[4]]
  ## ----------------------------------------------------------
  ## Create the data lists/rank lists based on miss. patterns:
  ## ----------------------------------------------------------
  YS <- Ys <- HYS <- HYs <- list()
  for(i in 1:a){temp.S <- temp.s <- list()
                for(r in 1:a)temp.S[[r]] <- fS(Y,d,i,r,a,Q)
                for(r in 2:a)temp.s[[r]] <- fs(Y,d,i,r,Q)
                YS[[i]] <- temp.S; Ys[[i]] <- temp.s
                for(r in 1:a)temp.S[[r]] <- fS(HY,d,i,r,a,Q)
                for(r in 2:a)temp.s[[r]] <- fs(HY,d,i,r,Q)
                HYS[[i]] <- temp.S; HYs[[i]] <- temp.s
  }
  ## ---------------------------------------------
  nYS <- lapply(YS,function(v)lapply(v,nrow))
  nYs <- lapply(Ys,function(v)lapply(v,nrow))
  nobs <- sum(d)
  n <- nrow(Y) ## no {NA,...,NA}- cases in Y (removed in smd.fun)

  ##----------------------#
  ##    Test Statistic    #
  ##----------------------#
  ##-------------------------------------------------#
  ## ehc.fun - Returns the vector of E(H(Yi)|Yr, S^i_r)#
  ##         - x = c(i,r)                             #
  ##-------------------------------------------------#
  ehc.fun <- function(x){
    i <- x[1]; r <- x[2]
    n01 <- nYS[[i]][[r]]
    n11 <- nYs[[i]][[r]]
    result <- numeric(0)
    if(n01>0 & n11>0){
      temp <-  cbind(rep(HYs[[i]][[r]][,1],n01),
                     rep(Ys[[i]][[r]][,2],n01),
                     rep(YS[[i]][[r]],rep(n11,n01)))
      k <- as.matrix(K(temp[,2]-temp[,3],b))
```

```
    sum.k <- rowsum(k,rep(1:n01,rep(n11,n01)))  # lenght= n01
    sum.k <- rep(sum.k,rep(n11,n01))
    w.vec <-k/sum.k
    result <- rowsum(temp[,1]*w.vec,rep(1:n01,rep(n11,n01))) #length = n01
  }
  result
}
ehc <- lapply(ir,function(v)lapply(v,ehc.fun))

EH <- list()
for(i in 1:a) {
  temp <- HYS[[i]][[1]][,1]
  for(q in 1:(a-1)) temp <- c(temp,ehc[[i]][[q]])
  EH[[i]] <- temp
}
HdF <- as.numeric(lapply(EH,mean))

##------------------------------------------#
## COVARIANCE CALCULATION - get the h(z)'s  #
##------------------------------------------#
##---------------------------------------------------------#
##  eH1.fun:  Returns E(H(Yi)|Ys=x,di=0,...,di_r=1,...)    #
##          - single value x                               #
##          - !!! r > 1 !!!                                #
##---------------------------------------------------------#
eH1.fun <- function(x,i,r){
  result <- numeric(0)
  if(length(Ys[[i]][[r]])>0){
    temp <- K(x-Ys[[i]][[r]][,2],b)
    temp <- temp/sum(temp)
    result <- HYs[[i]][[r]][,1]%*%temp
  }
  result
}

##---------------------------------------------------------#
##  Ht.fun:  Returns the vector of Ht.r^i(Zk,Zl)           #
##           for Yl in Ys(i,r) , constant k and  r > 1.    #
##---------------------------------------------------------#
Ht.fun <- function(p){
  i <- p[1] ; r <- p[2]
  n01 <- nYS[[i]][[r]]
  n11 <- nYs[[i]][[r]]
```

```
  result <- numeric(0)
  if(n01>0 & n11>0){
    A <- (HYs[[i]][[r]][,1]
          -apply(as.matrix(Ys[[i]][[r]][,2]),1,eH1.fun,i=i,r=r))
    result <- apply(YS[[i]][[r]],1,
                    function(x){
                      temp <- K(x-Ys[[i]][[r]][,2],b)
                      temp <- temp/sum(temp) #mean...
                      c(temp*A,0)
                    }) #ADD ZERO FOR DIMENTION PURPOSES
  }
  result
}

Ht <- lapply(ir,function(v)lapply(v,Ht.fun))

## Create H11(Z) vector to be used to calculate h(Z) vector.
## Arrange the values of H11(zj) in a vector matching the EH vector.
## for Di=1, H11(Z) splits up in the cases (11_), (101) & (100)
##  (i.e. in cases Ys(i,1),...,Ys(i,a),Y[di=1, dr=0 for r>1]).
##  -for Ys_type cases, H11(Zj)=sum_k(Htilta(Zk,Zj)) (vector tmp1)
##  -for Y[di=1, dr=0 for r>1], H11(Z)=0
## for Di=0, are the YS-type cases. (vector tmp2)

H11 <- matrix(rep(0,a*n),n,a)
for(i in 1:a){
  tmp1 <- tmp2 <- numeric(0)
  dd <- d[,i]
  for (r in 2:a){
    dd <- dd*(1-d[,Q[i,r]])
    tmpr <- rep(0,nYs[[i]][[r]])
    if(length(YS[[i]][[r]])>0 & length(Ys[[i]][[r]])>0){
      tmpr <- apply(Ht[[i]][[r-1]],1,sum)
      tmpr <- tmpr[-length(tmpr)]   # to remove the dummy 0 zero.
      ## tmp2 <- c(tmp2,apply(Ht[[i]][[r-1]],2,sum))
    }
    tmp1 <- c(tmp1,tmpr)
  }
  H11[1:length(tmp1),i] <- tmp1
  ##H11[,i]<- c(tmp1,rep(0,sum(dd)),tmp2) ## theoretically this is =0.
}

## ################################################################
```

```
  ## Create the h.mat matrix, using the EH list and H11 matrix    #
  ## ################################################################
  h.mat <- matrix(rep(0,a*n),n,a)
  for(i in 1:a) h.mat[,i] <- EH[[i]]-mean(EH[[i]])+H11[,i]

  pos <- 1:n
  track <- numeric()
  for(i in 1:a){
    temp <- pos[(d[,i]*d[,Q[i,2]]==1)]
    if (a > 2 )
      for(r1 in 3:a)temp <- c(temp,pos[(apply(cbind(d[,i],(1-d[,Q[i,2:(r1-1)]]),
                                            d[,Q[i,r1]]),1,sum)==r1)])
    temp <- c (temp,pos[(apply(cbind(d[,i],(1-d[,Q[i,2:a]])),1,sum)==a)])
    for ( r in 2:a )
      temp <- c(temp,pos[(apply(cbind((1-d[,Q[i,1:(r-1)]]),d[,Q[i,r]]),
                              1,sum)==r)])
    track <- cbind(track, temp)
  }

  h.mat1 <- matrix(rep(0,n*a),n,a)
  for(i in 1:a) h.mat1[track[,i],i] <- h.mat[,i]
  h.mat <- h.mat1

  ## N is deleted with N in the quadratic form, n=ni and h=hi.
  V <- (1/n)*cov(h.mat)  # n in this case is equal to n_i
  result <- list(HdF,V)
  result
}

##--------------------------------------------------------------------------#
## NP.Test Fucntion: Returns the proportions of rejection for alpha = 1% and 5% #
##                                                                          #
## Arguments: a= number of time points                                      #
##            N.vec = (N1,...,NI)                                           #
##            Q = matrix of j_r indices                                     #
##            smd.fun = function for simulation missng data                 #
##            b = vector of bandwidths in each group                        #
##            M = number of simultion runs                                  #
##            kernel = 0 -> Normal, 1-> Epanechnikov                        #
##--------------------------------------------------------------------------#

NP.Test <- function(a,N.vec,Q,smd.fun,b,M,kernel){
  if(missing(kernel)) kernel <- 0
```

```
## Create indices combination list to be used as a function argument.
## Use list form to avoid binding same length vectors into a matrix in ehc list.
ir <-list()
for(i in 1:a){
  temp <- list()
  for(r in 1:(a-1))temp[[r]]<- c(i,r+1)
  ir[[i]] <- temp
}
I <- length(N.vec)
if(length(b) ==1) b <- rep(b,I)   # for same b in everyt group.
## ------------------
## Contrast Matrices:
## ------------------
## F should be of the form: (F11,...,FIa)
C.ls <- list()
C.ls[[1]] <- kronecker(diag(rep(1,I)),M.fun(a))      # simple Time effect
if(I>1){
  C.ls[[2]] <- kronecker(M.fun(I),diag(rep(1,a)))    # simple Group effect
  C.ls[[3]] <- kronecker(M.fun(I),M.fun(a))          # Group by Time effect
}
Cmain.ls <- list()
Cmain.ls[[1]] <- kronecker(t(rep(1/I,I)),M.fun(a))   # main Time effect
if(I>1){
  Cmain.ls[[2]] <- kronecker(M.fun(I),t(rep(1/a,a))) # main Group effect
  Cmain.ls[[3]] <- kronecker(M.fun(I),M.fun(a))      # Group by Time effect
}
C.ls <- Cmain.ls                         # comment out for simple effects.

df.vec <-numeric()
for(l in 1:length(C.ls)){
    df.vec[l] <- nrow(C.ls[[l]])
}

TS.mat <- numeric()         # an Mx(number of tests) matrix of the TS-values.

## ----------------------- M- Simulation Runs ---------------------------
for(m in 1:M){
  data.list <- smd.fun(a,N.vec)
  Y.temp <- NULL; d.temp <- NULL; n.vec <- numeric()
  for (i in 1:I){
    Y.temp <- rbind(Y.temp,data.list[[i]][[1]])
    d.temp <- rbind(d.temp,data.list[[i]][[2]])
    n.vec <- c(n.vec,nrow(data.list[[i]][[1]]))
```

```
    }
    HY.temp <- matrix(as.vector(rank(Y.temp)),sum(n.vec),a)-1/2
    HY.temp[HY.temp > sum(d.temp)] <- NA
    HY.temp <- HY.temp/sum(d.temp)
    data.list[[1]][[3]] <- HY.temp[1:n.vec[1],]
    data.list[[1]][[4]] <- b[1]
    if(I>1)for(i in 2:I){
      data.list[[i]][[3]] <- HY.temp[(sum(n.vec[1:i-1])+1):sum(n.vec[1:i]),]
      data.list[[i]][[4]] <- b[i]
    }
    estimates.list <- lapply(data.list,estimation.function,a=a,Q=Q,ir=ir,
                             kernel=kernel)
    HdF <- numeric()
    V <- matrix(rep(0,(a*I)^2),a*I,a*I)
    for(i in 1:I){
      HdF[((i-1)*a+1):(i*a)]<- estimates.list[[i]][[1]]
      V[((i-1)*a+1):(i*a),((i-1)*a+1):(i*a)] <- estimates.list[[i]][[2]]
    }
    ## N in the enumarator is deleted with the N in the V-matrix.
    TS.mat <- rbind(TS.mat,sapply(C.ls,function(X)
                                  t(X%*%HdF)%*%solve(X%*%V%*%t(X))%*%X%*%HdF))
  }
  ## ----------------------------------------------------------------------------
  PR <- numeric()
  for(l in 1:length(C.ls)){
    PR <-  rbind(PR,c(sum((TS.mat[,l]>=qchisq(0.95,df.vec[l])))/M,
                      sum((TS.mat[,l]>=qchisq(0.90,df.vec[l])))/M))
  }
  TS.info <- t(apply(TS.mat,2,function(x)c(mean(x),var(x))))
  if(I == 1){
    dimnames(PR) <- list(c("T"), c("alpha=0.05","0.1"))
    dimnames(TS.info) <- list(c("T"), c("Mean(TS)" ," Var(TS)"))
  }
  if(I >1){
    dimnames(PR) <- list(c("T","G","TxG"), c("alpha=0.05","0.1"))
    dimnames(TS.info) <- list(c("T","G","TxG"), c("Mean(TS)" ," Var(TS)"))
  }
  result <- list(PR,TS.info)
  result[[1]][,2]
}
```

A.3 Code for ANCOVA Simulations

```
############################
## X= unif, sd(errors) =1 ##
############################
ws <- 1:12
B <- 5000
sizes30 <- c(30,30,30)
sdv <- 1

x.function <- function(sizes)runif(sum(sizes))
y.function <- function(x,mx,group,sdv){
  my <- apply(cbind(group,x),1,
           function(v)v[2]/mx *(v[1]==1)+(2-v[2]/mx)*(v[1]==2)+1*(v[1]==3))
  my+rnorm(length(group),0,sdv)
}

ANCOVA <- function(sizes,ws,B,sdv){
  N <- sum(sizes)      # sample size
  k <- length(sizes)   # number of groups
  group <- rep(1:k,sizes)        # (1,...,1,2,...,2,...,k,...k) 1XN
  allsizes <- rep(sizes,sizes)   # (n1,...,n1,...,nk,...,nk)    1xN

  ## Create an NxN matrix, each column=the rank of the ordered Xi's within group.
  ## each column out of N is equal to : (1,...,n1,1,...,n2,...,1,...,nk)'
  ## rij.mat is the same for all the simulation runs...get it outside for-loop.
  rij.mat <-NULL
  for( i in 1:k){
    rij.mat <-c(rij.mat,1:sizes[i])
  }
  rij.mat <- matrix(rij.mat,N,N)

  if (k>2) ctr.mat <- cbind(rep(1,k-1),diag(rep(-1,k-1)))
  if (k==2) ctr.mat <- matrix(c(1,-1),1,2)                 # C=(1, -1)

  PR <- NULL
  ##--------------------------------------------------------------#
  ##             First FOR-LOOP, different values of n             #
  ##--------------------------------------------------------------#
  for(m in ws){
    n <- m*2+1        # window size
    Q <-numeric()     # a vector to store the values of the Q-form
```

```
##-----------------------------------------------------------------#
##                    Second FOR-LOOP, B-runs                      #
##-----------------------------------------------------------------#
for(b in 1:B){
  x <- x.function(sizes)
  mx <- mean(x)
  ## Sort the data within each group, based on x.
  temp <- cbind(group,x)
  temp2 <- NULL
  for( i in 1:k){
    temp1 <- temp[temp[,1]==i,]
    temp2 <- rbind(temp2,temp1[sort.list(temp1[,2]),])
  }
  x <- temp2[,2]

  ## Calculate the rank of X <- r among each group, ni* G <- i(X <- r) ;
  x.mat <- t(matrix(x,N,N)) # each row is x'=(x11,...,xknk)=(x1,...,xN)
  temp <- (x.mat-x)
        # the columns are:(x1-x), ...(xN-x), i.e.all the dist.
  temp <- sign(temp)
  temp <- replace(temp, temp==0,1)
  temp <- (temp+1)/2
        # "-" =0 for xij>xr and "+"=1 for xij<=xr
  temp <- rowsum(temp,group)
        # ni*G <- i(X <- r), dim=kxN, "ranks within each group"
  ## correction for the two edges
  temp <- replace(temp, temp<n/2,(n+1)/2)
  temp <- sizes-temp+1
  temp <- replace(temp, temp<n/2,(n+1)/2)
  temp <- sizes-temp+1
  rr.mat <- apply(temp,2,rep,times=sizes) # NxN matrix
  ## K.mat
  k.mat <-1*(2*abs(rij.mat-rr.mat)/(n-1)<=1)

  y <- y.function(x,mx,group,sdv)

  ##-----------------------------------------------------------------#
  ## Calculation of the test statistic;
  ##-----------------------------------------------------------------#

  ## Obtain the Zrij's and the test statistic.
  Zrij.mat <- k.mat*y
```

```
         # an NxN matrix, in each column only k*n elemnts != 0
Zrit.mat <- matrix(Zrij.mat[!(Zrij.mat==0)],k*n,N)
         # remove 0's...get an (n*k)xN matrix.
group1 <- rep(1:k,rep(n,k))           # (1,...,1,....,k,...,k) 1x(n*k)
Zri.mat <- rowsum(Zrit.mat,group1)/n  # kxN, {\bZri.} r=1,...,N, i=1,...,k
Zi.vec <- apply(Zri.mat,1,mean)       # kx1, {\bZ.i.} i=1,...,k

##-----------------------------------------------------------------#
## calculate \hat \sigma <- {1,i}^2 and
##           \hat \sigma <- {1,i <- 1,i <- 2},i=1,...,k
##-----------------------------------------------------------------#

v1.mat<- var(t(Zri.mat))     #*(N-1)/N   # for s-plus var=sd^2

##-----------------------------------------------------------------#
## calculate  ci = temp1,i + temp2,i , i=1,...,k
##-----------------------------------------------------------------#
temp <- Zrit.mat[1:n,1:sizes[1]]
for(qq in 2:k){
  temp <- cbind(temp,Zrit.mat[((qq-1)*n)+(1:n),dim(temp)[2]+(1:sizes[qq])])
}
temp <- t(temp)
sigma.x <- apply(temp,1,var)  #(n-1)/n* ... sigma <- i^2(Xij), Nx1 vector
uij.vec <-(apply(k.mat,1,mean)/n )^2  # g/gi term, Nx1
temp1 <- sigma.x*uij.vec
temp1 <- N*rowsum(temp1,group)
#v2.mat <- diag(temp1)

temp2 <- k.mat
temp2[1:sizes[1],1:sizes[1]]=0
for(qq in 2:k){
  tmp1 <- sum(sizes[1:qq-1])+1
  tmp2 <- sum(sizes[1:qq])
  temp2[tmp1:tmp2,tmp1:tmp2]
}
temp2 <- apply(temp2,1,sum)
temp2 <- temp2*sigma.x
temp2<- rowsum(temp2,group)/(n^2*N)

##-----------------------------------------------------------------#
## calculate the estimate of the covariance matrix
##-----------------------------------------------------------------#
```

```
      V <-diag(temp1+temp2)  # for fixed X, V= diag(c1,...,ck)

      V <- v1.mat + V  # for random X (comment out for ficxed)

      ##-------------------------------------------------------------#
      ## calculate the statistic
      ##-------------------------------------------------------------#
      ta <- ctr.mat%*%Zi.vec                              # get CT
      cvc.mat <- solve(ctr.mat%*%V%*%t(ctr.mat))     # get [CVC']^-1

      ## QF - store the value in each run
      Q[b]<-N*t(ta)%*%cvc.mat%*%ta

    }  #for B

    ## Probabilities of rejection
    temp <- cbind(table(Q>=qchisq(0.90,k-1))[2]/B,
               table(Q>=qchisq(0.95,k-1))[2]/B,
               table(Q>=qchisq(0.99,k-1))[2]/B)
    PR <- rbind(PR,temp)

  } # for m in const

  PR <- cbind(2*ws+1,PR)
  dimnames(PR) <- list(NULL,c("n","a=0.1","a=0.05","a=0.01"))

  result <- list("Sample sizes, Number of sim.runs, seeds"
                 =c("n"=sizes,"B"=B),
                 "Prob. of Rejection (alpha=.1, .05 & .01)"=PR)
  result
 }

#################################################################
temp30 <- ANCOVA(sizes30,ws,B,sdv)
temp30
```

Bibliography

[1] Aerts, M., Claeskens, G., Hens, N. and Molenberghs, G. (2002). Local multiple imputation. *Biometrica*, 89, 375–388.

[2] Akritas, M. A. Antoniou, E. S. and Wang, L. (2004). Fully Nonparametric ANCOVA with Fixed Window Size. *submited.*

[3] Akritas, M. A. and Arnold, S. F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs. *Journal of the American Statistical Association*, 89, 336–343.

[4] Akritas, M. A. and Arnold, S. F. (2000). Asymtotics for Analysis of Variance When the Number of Levels is Large. *Journal of the American Statistical Association*, 95, 212–226.

[5] Akritas, M. G, Arnold, S. F. and Brunner, E. (1997) Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs. *J. American Statistical Association* 92, 258-265.

[6] Akritas, M. G, Arnold, S. F. and Du, Y. (2000). Nonparametric models and methods for nonlinear analysis of covariance. *Biometrika.* 87, 507-526.

[7] Akritas, M. G., and Brunner, E. (1997). A Unified Approach to Rank Tests for Mixed Models. *J. Statist. Plan. Infer*, 61, 249–277.

[8] Akritas, M. G., Osgood, D. W. and Kuha, J. (2002). Nonparametric Approach to Matched Pairs Missing Data, *Sociology Methods* & *Research*, 30, 425–454.

[9] Akritas, M. G. and Wang, L. (2002). Lack-of-fit test in nonparametric random design regression. *submitted*.

[10] Antoniou, E., S., Akritas, M. G. and Kuha, J. (2004). Data Analysis under MCR Missingness via Nonparametric Imputations. *submitted*.

[11] Arnold,S. F. (1980). Asymptotic validity of F-tests for the ordinary linear model and the multiple correlation model. *Journal of the American Statistical Association*, 75, 890–895.

[12] Arnold,S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

[13] Bates, D.M. and Pinheiro, J.C. (1998). Computational methods for multivariate level models. *Technical Memoradum BL0112140-980226-01 TM*, Bell Labs, Lucent Technologies, Murray Hill, NJ.

[14] Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist.* 12, 1034–1039.

[15] Brunner, E., Dette, H. and Munk, A. (1997). Box-type approximations in nonparametric factorial design. *Journal of the American Statistical Association*, 92, 1494–1501.

[16] Brunner, E., Munzel, U. and Puri, M., L. (1999). Rank-Score Tests in Factorial Designs with Repeated Measures. *Journal of Multivariate Analysis*, 70, 286–317.

[17] Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 89, 81–87.

[18] Cheng, P. E. and Chu,C. K. (1996). Kernel estimation of distribution functions and quantiles with missing data. *Statistica Sinica* 6, 63–78.

[19] Collins, M, G, Schafer, J, L and Kam, C. (2001). A Comparison of Inclusive and Restrictive Strageties in Modern Missing Data Procedures. *Psycological Methods* 6, 330–351.

[20] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via EM algorythm. *Journal of the Royal Statistical Society Series B* 39, 1–38.

[21] Efron, B. (1994). Missing data imputation, and the bootstrap. *Journal of the American Statistical Association* 89, 463–475.

[22] Graham, J. W, and Hofer, S. M (1993). *EMCOV.EXE user's guide.* Alhambra, CA: University of Southern California, Institute for Prevention Research.

[23] Gold, M. and Osgood, D. W. (1992). *Personality and Peer Influence in Juvenile Corrections*. Westlake, CT: Greenwood.

[24] Hewitt, E. and Stromberg, K. (1969). *Real and Abstract Analysis*, 2nd ed. Springer, Berlin.

[25] Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 28, 963–974.

[26] Lehmann, E. L. (1998). *Elements of Large-Sample Theory*. Springer-Verlag, New York.

[27] Li, K. H., Meng, X. L, Raghunathan, T, E and Rubin, D, B (1991) *Significance levels from repeated p-values with multiply-imputed data Statistica Sinica*, 1, 65–92.

[28] Li, K. H., Raghunathan, T, E and Rubin, D, B (1991) *Large Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and an F Refence Distribution Journal of the American Statistical Association*, 86, 1065–1073.

[29] Little, R.J.A. (1995). Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association*, 90, 1112–1121.

[30] Little, R. J. A., Milliken, G., A, Stroup, W, W, and Wolfinger, R., D. (1996). *SAS system for mixed models*. SAS Institute, Cary, New York.

[31] Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

[32] Norusis, M. J. (2000). *SPSS 10.0 guide to data analysis.* Prentice Hall, Upper Saddle River, NJ.

[33] Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects Models in S and S-PLUS*. Springer, New York.

[34] Pyke, R. (1965). Spacings (with discussions). *Journal of the Royal Statistical Society. Series B.* 27, 395–449.

[35] Rosenbaum, P., R, and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causial Effects. *Biometrika*, 70, 41–55.

[36] Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.

[37] Rubin, D.B. (1987). *Multible imputation for nonresponse in surveys.* Wiley, New York.

[38] Sacks, J. and Ylvisaker,D (1970). Designs for regression problems with correlated errors III. *Ann. Math. Statist.* 41, 2057–2074.

[39] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London.

[40] Schafer, J. L. (1999). *Multiple Imputation of Incomplete Multivariate Data under a Normal Model*. Softaware for Windows. Penn State University, University Park, PA.

[41] Schafer, J. L. (2001). *Multiple Imputation with PAN.*

[42] Serfling, R.J. (1980), *Approximation theorems of mathematical statistics*. Wiley, New York.

[43] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, 6, 177–184.

[44] Schimert, J., Schafer, J. L., Hesterberg, T., Fraley, C. and Clarkson, D. (2001), *Analysing Missing Values in S-plus*. Insightful Corporation, Seatle, WA.

[45] Tang, G., Little, R. J. A. and Raghunathan, E., T. (2003) Analysis of Multivariate missing data with nonignorable nonresponce. *Biometrica*, 90, 747–764.

[46] Titterington, D. M. and Mill G. M. (1983). Kernel-based density estimates from incomplete data. *Journal of the Royal Statistical Society B,* 45, 258–266.

[47] Titterington, D. M. and Sedransk J. (1989). Imputation of missing values using density estimation. *Statistics & Probability Letters,* 8, 411–418.

[48] Tsangari, H. and Akritas, M. G. (2003). Nonparametric ANCOVA with two and three covariates. *Journal of Multivariate Analysis,* in press.

[49] Uspensky, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.

[50] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall, London.

[51] Wang, H. and Akritas, M. G. (2004). Inference from Heteroscedastic Functional data, Part I: Identical Sampling Points, *in preparation*.

[52] Wang, L. and Akritas, M. G. (2002). Two-way heteroscedastic ANOVA when the number of levels is large, *submitted*.

Vita

Efi Antoniou

Personal

- Birthdate: February 3, 1978
- Citizenship: Cypriot

Education

- Ph.D. in Statistics, The Pennsylvania State University, August 2004
- M.S. in Statistics, The Pennsylvania State University, December 2001
- B.S. in Mathematics and Statistics, University of Cyprus, June 1999

Academic Experience at the Pennsylvania State University

- 1999–2003 Teaching Assistant
- 2001–2004 Research Assistant
- 2001 Graduate Consultant

Professional Affiliations

- The Institute of Mathematical Statistics